Using natural variation and adaptive evolution of Metschnikowia pulcherrima for lipid production from low value plant feedstocks

Hicks, Robert

Award date:
2019

Awarding institution:
University of Bath

Using natural variation and adaptive evolution of *Metschnikowia pulcherrima* for lipid production from low value plant feedstocks

Robert Henry Hicks

A thesis submitted for the degree of Doctor of Philosophy University of Bath

Department of Biology and Biochemistry

January 2019

Table of Contents

Acknowledgements

I would like to begin by thanking my lead supervisor, Daniel Henk, who has been an open and engaged source of guidance and help throughout my PhD, particularly with the latter bioinformatics aspects. Daniel has also been a good friend, as well as excellent BBQ host; never has a vegetarian cooked a better burger. I would also like to thank my co-supervisor, Chris Chuck, for his advice and enthusiasm throughout all projects within my PhD. I would also like to thank Chris for being the driver behind the wider *Metschnikowia pulcherrima* project. It was important for me to work on something which had genuine industrial relevance, so this project has been perfect.

Next, I would like to thank and acknowledge those I have worked with, beginning with fellow Henk group members Debs and Mauro, who are both good friends and lovely people to share a lab with. I am looking forward to working closer with (or in Debs' case, working for) and learning from you both in the coming months. Apologies in advance for some incoming ignorance. Thanks also go to Felix for being a great link between Chem Eng and Biology, a helpful HPLC hand, and a great person to share a car with for the four-hour journey to Aberystwyth. Thanks to Yuxin and Jenny who worked under my supervision and contributed to aspects of this thesis. Thanks to David Leak, Rod Scott, Chris Pudney and Jim Caunt for helpful discussions along the way about my work, or future ideas. Thanks also to members of the wider project: Sophie, Haddy and Luca.

I would also like to acknowledge all the friends I have made over the course of my PhD. A special 'GBS' shout out goes to the close friends I have made from my CSCT cohort. Many, many memories have been made, and fortunately/unfortunately, some of them are documented. I would also like to acknowledge friends I have made outside of the University, who have made living in Bath a really enjoyable experience and have given me great support. I'd also like to acknowledge the people I met during my time with Rainforest Foundation Norway who helped me to fully appreciate the 'bigger picture' of my project.

The final thanks goes to my family. Firstly, to my sister, Carly, who has always been available for career/life advice… and the not-so-occasional grammar question or CV proof read. Secondly, to Mum and Dad - from the start of my undergraduate degree to the end of my PhD (including my brief foray into employment between the two), you have given me all the help and support I have ever needed. This is epitomised by the 13 (I think) occasions where you have helped me to move in/out of various houses/dorms/flats, for various reasons, over that period (including a special mention to *that* sofa). I would like to say there won't be many more moves, but why break the habit of a lifetime. Thanks for everything.

Abstract

The oleaginous yeast *Metshnikowia pulcherrima* has previously been investigated as a potential platform organism for microbial oil production. This is due to several key phenotypes in addition to efficient oil production, including its ability to be grown in non-sterile conditions whilst metabolising a range of oligo- and monosaccharide carbon sources within lignocellulosic hydrolysates. To further improve the industrial potential of this organism in the absence of an appropriate genetic toolkit, adaptive laboratory evolution was applied to selected strains.

First, strains were adapted to lignocellulosic fermentation inhibitors in two strategies; either a single inhibitor, formic acid, or an inhibitor cocktail containing formic and acetic acid, 5-HMF and furfural. Phenotypic analysis of evolved cell lines reveals improved tolerance versus the progenitor across all strains, including cross adaptation of single-inhibitor evolved strains to conditions where all four inhibitors are present. Interestingly, the lipid production of the inhibitor cocktail evolved strains markedly increases, with one strain, named '4x3', producing 41% lipid by dry weight compared to 22% by the progenitor. Following this, 4x3 was taken forward for further characterisation within 2L bioreactors and whole genome sequencing was performed to prospect for causative genetic modifications underpinning improved inhibitor tolerance and lipid production.

Strain 4x3 was then subject to a second round of adaptive laboratory evolution to improve its poor utilisation of xylose. Here, though all evolved cell lines had improved xylose metabolism versus the progenitor, a large degree of variation was observed suggesting multiple evolutionary routes to phenotypic improvement. One cell line in particular showed a drastic improvement, producing over 17x the biomass of the progenitor when grown on xylose as the sole carbon source. Whole genome sequencing was again performed on all strains, revealing both highly modified and dissimilar genomic structures, characterised by a high frequency of copy number variants. From this data, causative mutations are suggested and a general theory

explaining the novel genome complexity following adaptive laboratory evolution is presented and discussed.

To support this study and the wider *M. pulcherrima* project, a simple, image based high-throughput lipid quantification method was developed after the commonly applied method of lipid straining was found to be inconsistent for this organism. The developed method allows for population dynamics to be tracked which was applied to overcome a potential production bottleneck within a semi-continuous bioprocess.

Contributions

Contributions for chapters within this thesis submitted as publications/manuscripts prepared for submission are given within the declaration of authorship at the end of each relevant chapter.

There are no further collaborations outside of those stated above.

References for prepared manuscripts are placed within each relevant chapter. References for sections not submitted for publication or prepared for submission are collected at the end of this thesis.

Abbreviations.

Relevant abbreviations are provided within the text for submitted manuscripts or those in preparation, however the following is a list of those used throughout this thesis.

5-HMF – Hydroxymethylfurfural
ALE – Adaptive Laboratory Evolution
CNV(s) – Copy Number Variant(s)
DMSO - Dimethyl sulfoxide
FAME – Fatty Acid Methyl Ester
PPP – Pentose Phosphate Pathway
SNP(s) – Single Nucleotide Polymorphism(s)

1. Literature Review

1.1. Sustainability of Vegetable Oils

The rising popularity of vegetable derived oils for use as biodiesels, oleochemicals and components within food and cosmetics has intensified the expansion of large-scale dedicated plantations. Among the popular vegetable crops for oil production, oil palm is currently the world's main contributor due to an unrivalled per hectare yield, access to low labour costs and unique chemical properties which make it well suited for use within food[1,2]. As oil palm thrives in tropical climates, 80% of global production comes from Indonesia and Malaysia where yields benefit from year-round consistency in temperature and rainfall[3]. As demand for palm oil has increased due to a shifting trend away from fossil derived chemicals, the expansion of monoculture palm plantations have caused widespread displacement of primary rainforests within tropical regions. In Indonesia alone, an estimated 0.84 million hectares of primary forest was lost every year between 2002 and 2012 with palm oil plantation expansion responsible for over half of this[3]. As expansion has occurred there has been an inevitable loss of biodiversity. It is reported for example that a hectare of rainforest can contain over 200 plant species and that 60% of the total plant and animal species present are endemic only to these regions[3]. Furthermore, the presence of palm plantations can directly affect nearby forests through habitat fragmentation, proximity pollution and the introduction of non-native species such as rats, which negatively influence existing food chains[4]. An often-overlooked consequence of palm oil expansion is the conversion of tropical peatlands[5]. Until recently, peatlands have escaped plantation conversion due to the costly drainage procedure required before they're made suitable. With new technological developments however, these areas have become increasingly targeted for land conversion, leading to at least a third of new plantations coming from drained peatland[5]. Although the biodiversity of these areas is not at the same level as primary rainforest, peatlands are one of Earth's most effective mechanisms of biological carbon storage meaning that their conversion releases thousands of years of stored CO_2. Taking into account the combined displacement of rainforest and

peatlands, it is estimated that the overall emissions from palm derived biodiesel is up to three times higher than the fossil fuels they are brought in to replace[6].

Though the food industry remains the greatest consumer of palm oil globally, the rate at which demand will grow will be significantly influenced by policies surrounding Indonesia's domestic consumption and the global transportation sector. Indonesia, who in the past have almost exclusively acted as an exporter of palm oil, have recently proposed an aggressive expansion to domestic consumption. Through its biofuel blending mandate, by 2020 it will be required for traditional diesel to be blended to include 30% biodiesel (synonymous with palm biodiesel in Indonesia's case) for use within transportation[7]. This policy will be financially subsidised by a crude palm oil levy applied to palm oil exports, which will be fed back into plantation expansion[8]. The Renewable Energy Directive mandated by the European Union, in an attempt to lessen fossil fuel reliance and promote domestic vegetable oil production, has so far also had significant implications on palm oil consumption. From 2000 to 2014, biodiesel consumption and vegetable oil production within the EU have gone from effectively zero, to approximately 12 and 6 million tonnes respectively (Figure 1). This ~ 6 million tonne shortfall between consumption and production in 2014 for EU vegetable oil production has meant it necessary for imports to subsidise the demand, with 5.7 million tonnes of this coming directly from palm oil[7]. The magnitude of these imports have now made biodiesel overtake the food industry as the main consumer of palm oil within the European Union (Figure 2). At present, the EU is in the process revising the Renewable Energy Directive, and despite support within the EU commission to acknowledge and reassess the legislation permitting palm oil's inclusion within biodiesel[9], the latest draft contains no explicit reference to palm oil[10].

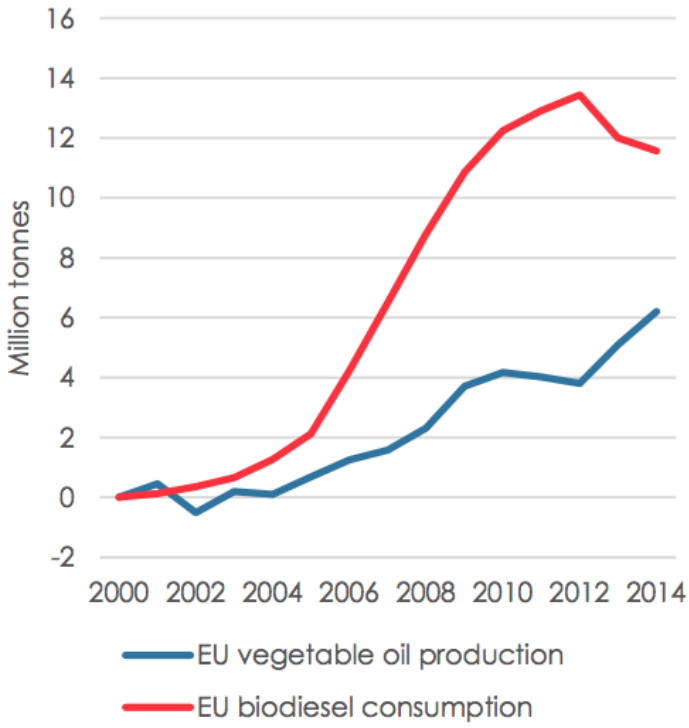


Figure 1. Biodiesel consumption compared to vegetable oil production within the EU. Source: 2018 report by Chris Malins - Driving Deforestation[7]. To match biodiesel demand, the EU has imported 5.7 million tonnes of palm oil in 2014.

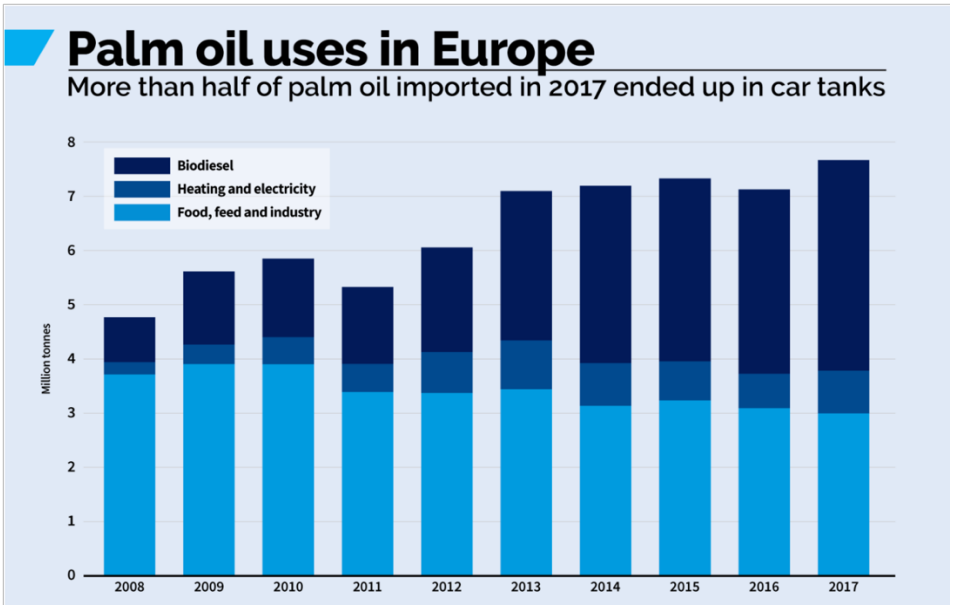


Figure 2. Changing use of palm oil imports within the European Union. Source: Transport & Environment – Seven Facts About Palm Oil Biodiesel[11].

1.2. Microbial oil

1.2.1. Oleaginous Yeasts

Oils from microorganisms have great potential to provide sustainable alternatives to fossil and vegetable fuels and bulk chemicals. Early commercial interest focused primarily on oleaginous microalgae, however issues relating to their application at industrial scale i.e. inherent low culturing densities, long culturing times, difficulties in processing the large volumes (dewatering), inconsistent growth due to seasonal variation and a predisposition to contamination by faster growing organisms, have meant progress has stalled with few large scale successes[12]. Oleaginous yeasts have since emerged as more realistic alternatives for large-scale production of bulk chemicals, outperforming microalgae in growth rate, production rate and cell density[13].

Oleaginous status is given to microorganisms when greater than 20% of their total dry weight is attributed to stored lipids. Within yeasts, just 5% of species are able to accumulate more than 25% lipid with species of *Rhodotorula, Yarrowia* and *Lipomyces* attracting the most attention within biofuel research[14][15]. Ecologically, oleaginous species are generally found in relatively dry, nutrient poor habitats, where the storage of lipids is advantageous and essential for survival[13]. Ideal parameters for lipid accumulation are generally nitrogen (and to a lesser extent phosphorus) limited, but carbon source excessive. In this scenario, nitrogen dependant mechanisms of protein and nucleic acid biosynthesis cease, causing growth rates to slow and available carbon to be re-routed towards lipid biosynthesis and storage within intracellular lipid droplets[16].

Lipids are stored within cells as lipid bodies or droplets as triacylglycerols (TAG); esters formed from glycerol linked to three fatty acids. The molecular composition of these fatty acids can vary within yeasts and fungi, but within *S. cerevisiae* oleic (C18:1), stearic (C18:0) and palmitic acid (C16:0) predominate. Within other yeasts, fatty acids such as linoleic, linolenic and palmitoleic acids are also synthesised[17]. As the fatty acid profile of yeasts is quite similar to the profile in plant derived oils, they

are a good replacement for these in sectors such as food and biodiesel[15]. FAME (fatty acid methyl ester) profiling is the technique used to characterise the fatty acid profile of oleaginous organisms. The FAME profile can be greatly affected by growth conditions, with factors such as carbon source, nitrogen and oxygen availability and age of culture all influencing total fatty acid abundance[18]. Fatty acid ratios and abundance is also varied within strains of the same species[16].

1.2.2. Lipid biosynthesis: oleaginous vs. non-oleaginous yeast.

Although the mechanism of lipid biosynthesis has been well described within the non-oleaginous *S. cerevisiae,* until recently a detailed comparative model within an oleaginous organism has not been available[19]. Now, given the influx of interest in industrial potential, *Yarrowia lipolytica* has emerged as a model oleaginous yeast. A fully sequenced genome and an ever developing genetic toolkit means that important genetic distinctions can now be made between an oleaginous and non-oleaginous yeast[20].

In oleaginous organisms, lipid biosynthesis is inherently linked to nitrogen availability. Upon nitrogen starvation, lipid accumulation is thought to be initially triggered by a decrease in intracellular adenosine monophosphate (AMP) concentrations. This in turn causes a downregulation of isocitrate dehydrogenase activity within the citric acid cycle, as within oleaginous yeast, this enzyme requires sufficient AMP to function[19]. In conditions where nitrogen is available, sufficient AMP enables isocitrate dehydrogenase to convert citrate to isocitrate, however upon nitrogen starvation, the AMP is instead used to generate ammonium for protein biosynthesis leading to a decrease in cellular concentrations. The inactivity of isocitrate dehydrogenase due to reduced AMP leads to an accumulation of citrate, a key molecule within lipid biosynthesis, within the mitochondria[19]. It is the accumulation of citrate which is one of the key distinctions between oleaginous and non-oleaginous yeasts, with it reported that mitochondrial citrate levels are between three and four times higher in lipid accumulating species, and the transporters responsible for relocating citrate to the cytoplasm are more efficient [14].

In both oleaginous and non-oleaginous organisms, acetyl-coenzyme A (acetyl-CoA) is a key molecule and building block in lipid biosynthesis[19]. Within *S. cerevisiae,* acetyl-CoA is generated enzymatically by acetyl-CoA synthase from acetate originating from glycolysis. Here, *S. cerevisiae* has two enzymes catalysing acetyl-CoA generation: *ACS1* and *ACS2*, though *ACS2* has more involvement within fatty acid biosynthesis[21]. In contrast, *Y. lipolytica* encodes only one acetyl-CoA synthase type enzyme, which has greater similarity to *S. cerevisiae ACS2*[19]. A key difference however is that within *Y. lipolytica* the majority of acetyl-CoA generation is derived from the cleavage of citrate by ATP citrate lyase (ACL). Therefore this enzyme is particularly active under nitrogen limitation when an accumulation of citrate occurs (Figure 3)[22].

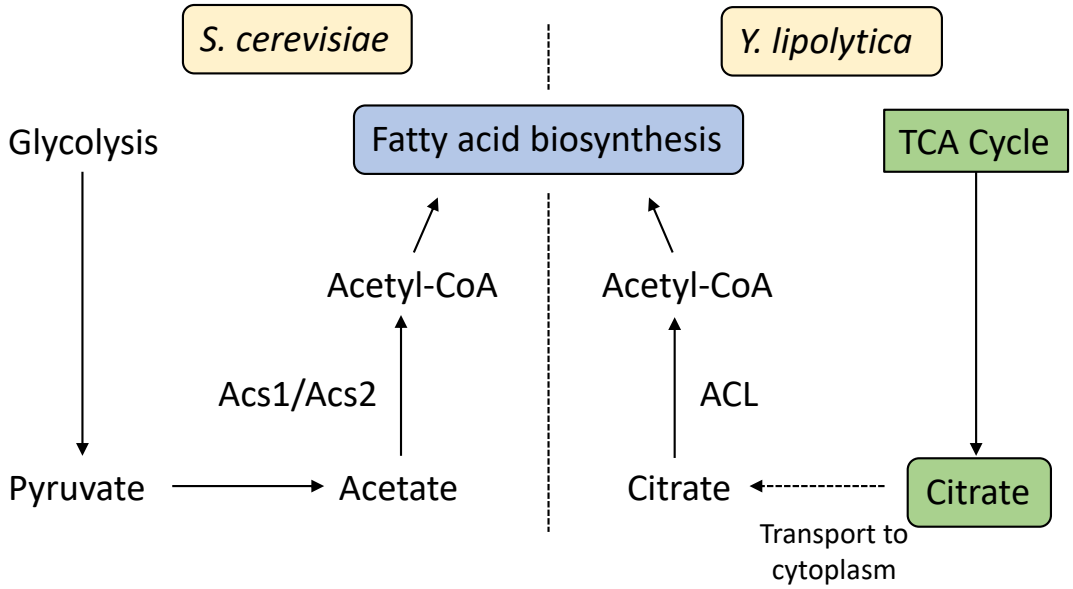


Figure 3. Generation of acetyl-CoA in non-oleaginous *S. cerevisiae* and oleaginous *Y. lipolytica*. In S. cerevisiae, pyruvate from glycolysis is converted acetate, which is then converted through the action of two acetyl-CoA synthase enzymes Acs1/Acs2 into acetyl-CoA. In oleaginous yeast, acetyl-CoA is generated from citrate from the TCA cycle by ATP citrate lyase (ACL). Citrate concentrations in oleaginous yeasts are elevated in response to nitrogen limitation. Steps in green occur within the mitochondria.

NADPH is also key within lipid biosynthesis pathways, with two required for each fatty acid elongation step and one required for desaturation of the molecule. Although is it known that oleaginous organisms possess comparatively large intracellular stocks of this cofactor compared with non-oleaginous, it was unknown

until recently where *Y. lipolytica* was sourcing this molecule[23]. Work within other oleaginous organisms found malic enzyme to be the generator of NADPH, as overexpression of this gene resulted in a 2.5-fold increase in lipid accumulation in *Mucor circinelloides*[24]. Malic enzyme within *Y. lipolytica* however has been shown localise within the mitochondria rather than the cytoplasm, and to prefer NAD^+ to $NADP^+$, suggesting it does not generate NADPH within fatty acid biosynthesis[23]. Furthermore, *Y. lipolytica* malic enzyme expression does not correlate with nitrogen limitation[25]. Instead, most recent evidence suggests that the oxidative pentose phosphate pathway (PPP) is responsible for NADPH generation in *Y. lipolytica*, with PPP-specific NADPH increasing two-fold during lipid biosynthesis[23].

Within *Y. lipolytica*, fatty acid synthesis takes place within the cytoplasm, and begins with the irreversible conversion of acetyl-CoA into malonyl-CoA by *ACC1*; a committing step within the pathway. Next, the fatty acid synthase (FAS) complex causes a condensation reaction between the malonyl-CoA generated previously and another acetyl-CoA, resulting in the addition of two carbons; the first link within the fatty acid chain. Repeated additions of acetyl-CoA leads to the elongation of the fatty acid chain, two carbons at a time, consuming two NADPH molecules per cycle[19]. The FAS complex releases fatty acids of C16:0 or C18:0 chain length, where they can be further modified by elongases and desaturases to make longer chains or desaturated chains (such as C16:1 or C18:2 respectively)[26]. Fatty acids are then assembled into TAG by esterification to a glycerol-3-phosphate backbone within the Kennedy pathway, causing the formation of lipid bodies (Figure 4)[27].

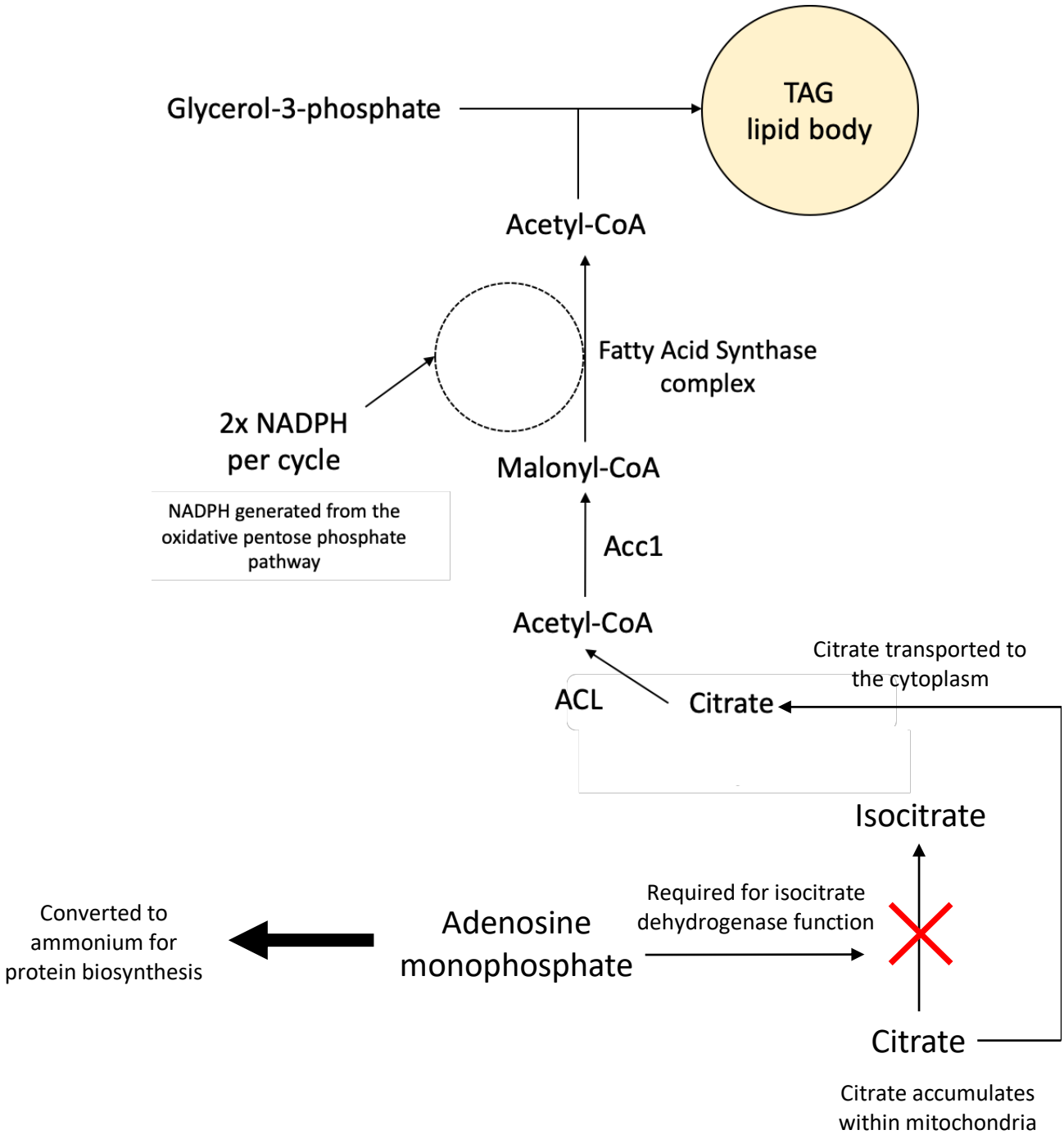


Figure 5. Summary of lipid accumulation in *Y. lipolytica*. Nitrogen limited conditions causes adenosine monophosphate (AMP) to be used to generate ammonium for protein biosynthesis. Due to low AMP levels, isocitrate dehydrogenase, which requires AMP, is unable to convert citrate to isocitrate, meaning citrate accumulates within the mitochondria. Citrate is transported into the cytoplasm, where it is converted to acetyl-CoA by ATP citrate lyase. Acc1 converts acetyl-CoA into malonyl-CoA, and successive acetyl-CoA's are added by the fatty acid synthase complex to elongate the fatty acid chain at the expense of two NADPH molecules per addition. NADPH in *Y. lipolytica* originates from the oxidative pentose phosphate pathway. The

elongated fatty acid is assembled into TAG by esterification to a glycerol-3-phosphate backbone to form a lipid body.

1.3. Second generation feedstocks

Technoeconomic analysis states that scaled microbial oil production is more expensive than vegetable oils when a glucose-based bioprocess is considered. Within this report, it estimated that the cost of a purified microbial oil within a process yielding 10,000 tonnes a year to be $3.4/kg, even if the cost of glucose is zero (assuming glucose was sourced through waste or by product streams)[28]. This price rises to $5.5/kg for oil and $5.9/kg for biodiesel when glucose is costed at $400/t. Sunflower and peanut oil, as a comparison, were priced at $1.6/kg and $2.1/kg respectively. Figures sourced from a similar time unsurprisingly price palm oil cheaper still at $770 per tonne or $0.77/kg[29]. As a cost-free glucose is an unlikely feedstock, it is clear from these figures that the only realistic chance of an economically viable alternative would require low-cost second generation feedstocks. In this way, it would be both cheap to source, and not be associated with the food versus fuel debate which would be landed on a glucose based process. To further improve the financial viability of lignocellulosic derived microbial oil, a biorefinery approach, where co-products are produced in addition to the bulk microbial oil, is thought to be required[30]. In oleaginous yeasts, one potential application of the biorefinery approach is to utilise the yeast itself as a component for use in animal feeds[31].

Second generation feedstocks are typically lignocellulosic based materials and include purpose-grown energy crops such as switch grass or miscanthus, agricultural wastes such as wheat straw or corn stover and municipal wastes including food waste or spent coffee grounds. Of these, the true waste feedstocks are the cheapest and available between $0-60 per tonne, whereas the dedicated energy crops are priced between $100-120 per tonne[32]. Though exact structural compositions of lignocellulose feedstocks differ from source to source, the same three constituents; cellulose, hemicellulose and lignin, are conserved. The cellulose component consists of β-(1,4) linked glucan chains, and is the most glucose rich and therefore of most

value to microbial growth. The hemicellulose component also contains a significant amount of carbon however these are predominantly in the form of branched five- and six carbon polysaccharides. The lignin component is of least value to a biological process, consisting of an aromatic three dimensional polymer of phenylpropane units which form a barrier around the cell wall polysaccharides[33]. The ratios of each component vary source to source, for example rice straw is composed of 7% lignin and 40% cellulose, compared to 29% and 33% respectively in sugar cane bagasse[32]. The ratios of the sugars can also vary, with wheat pulp containing 16% xylose compared to just 3.5% in pine pulp[34]. Though abundant in carbon, a main challenge of lignocellulosic bioprocessing comes from first extracting the sugars, and then whether the available carbon is able to be metabolised and converted into product by the potential platform organism. For instance, optimal growth on lignocellulosic feedstocks by the common industrial strain *S. cerevisiae* is limited as it cannot utilise the xylose components present. Therefore, potential platform organisms for lignocellulosic based bioprocesses are ideally able to metabolise all of the carbon sources available, or made suitable through genetic modification, as is well documented within *S. cerevisiae*[35].

1.3.1. Fermentation inhibitors from lignocellulosic feedstocks

Due to the tightly organised structure of this material, pretreatment is required to release the sugar from lignocellulose before fermentation can take place. Common methods of pretreatment include acid hydrolysis, acid assisted steam explosion, hydrothermal processing and mild alkaline treatment[36]. Pretreratment methods of all types are often followed by the addition of enzyme mixes in order to hydrolyse oligosaccharides not converted[37]. Naturally, pretreatment each method has its own merits and reasons for selection, for instance acid-based methods are effective in hydrolysing hemicellulose into monosaccharides, whereas mild alkaline methods are more effective in removing lignin[36]. Due to the intense conditions required, almost no pretreament method is capable of efficient sugar release without the formation of toxic by-products, termed lignocellulosic fermentation inhibitors. These compounds are deleterious to growth and product formation, meaning bioprocesses

are either forced to remove these compounds post treatment, or limit processes to organisms able of tolerating the presence of fermentation inhibitors. Pretreatment methods using ionic liquids have been shown as an effective way limit the formation of fermentation inhibitors from lignocellulosic biomass, however if even small amounts of ionic liquid remain in the treated material, this can be toxic to the microorganism[38].

Fermentation inhibitors can be classified according to their functional group; carboxylic acid, ketone, phenolic or aldehyde (Figure 6)[39]. Two of the most abundant inhibitors present in pretreatment methods are the aldehydes, consisting of 5-hydroxymethyl furfural (5-HMF) and 2-furaldehyde (furfural) and formed as a result of the dehydration under intense heat and pressure of hexose or pentose sugars respectively[40,41]. The most common inhibitory acids generated are formic and acetic acid, with formic acid produced as a derivative of furfural or 5-HMF and acetic acid produced when acetyl groups are released from hemicellulose[42]. Finally, the lignin present within hemicellulose is degraded into phenolic derivatives such as vanillin and catechol, which although are found at lower concentrations than 5-HMF or furfural, can be comparatively more inhibitory[43]. Lignin which is not degraded presents further downstream complications, as enzymes used during biological pretreatments such as cellulases are known to adsorb into lignin[44]. For typical examples of fermentation inhibitor concentrations released from lignocellulose pretreatment, see Table 1.

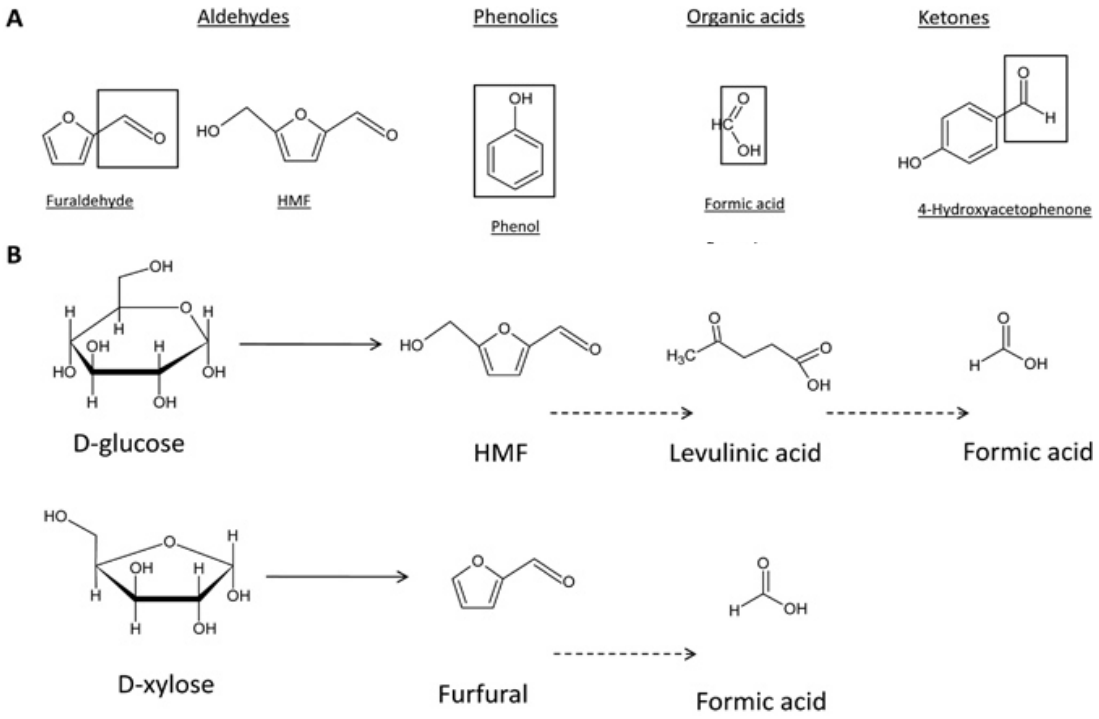


Figure 6. Common inhibitors present within pretreated feedstocks. A – Classification of inhibitor types, with the functional groups boxed. B – Formation route of the four most abundant inhibitors as break down products of D-glucose and D-xylose. Figure taken from Taylor[39].

Table 1. Lignocellulosic derived fermentation inhibitors. Pretreatments used in each case: Sugarcane – HCl hydrolysis, Wheatstraw – H_2SO_4 hydrolysis, Soft wood – dilute HCl hydrolyusis, Spruce – dilute H_2SO_4, Eucalyptus globulus - H_2SO_4 hydrolysis. Table adapted from Chandel[45]

Lignocellulosic Feedstock	Inhibitor profile (g/L)
Sugarcane Bagasse	Fufural, 1.98; Phenolics, 2.75; Acetic acid 5.45
Wheat Straw	Furfural, 0.15; Acetic acid 2.70
Soft Wood	Acetic acid, 5.3; Fufrural, 2.2
Spruce	Furfural, 1; 5-HMF, 3.3; Acetic acid, 5; Formic acid, 0.7; Levulinic acid, 0.2
Eucalyptus globulus	Furfural, 0.25; 5-HMF, 0.07; Acetic acid, 3.41; Phenolyics, 2.23.

The inhibitory mechanisms of these compounds are varied and complex. 5-HMF and furfural non-specifically react with RNA, DNA and proteins, as well as causing membrane damage resulting in disruptions to metabolism and cell viability[46]. Acetic and formic acid enters cells via diffusion where within the cytoplasm they dissociate

into acetate/formate and a proton. Proton accumulation leads to cytoplasm acidification, impairing metabolism by inhibiting key glycolytic enzymes and NADH dehydrogenases, leading to increased lag times and reduced growth[42]. Phenolic compounds also diffuse into the cells, where they alter membrane lipid/protein ratios causing increased fluidity and permeability, leading to cell leakage[47]. In addition to the individual effects of each inhibitor, studies investigating the effects of an inhibitor cocktail found that they act synergistically to create an effect far beyond the sum total of each individual effect[48,49].

The effects of fermentation inhibitors on growth and productivity has been investigated within many conventional and non-conventional yeasts, but only within *S. cerevisiae* have the response mechanisms which confer tolerance been thoroughly investigated. *S. cerevisiae* is comparatively tolerant to aldehyde inhibitors and the expression of three MAPK-signalling pathways and the phosphatidylinositol signalling pathway has been shown to be key in response to 5-HMF exposure[50]. Furthermore, proteins involved in oxidative and osmotic stress response, DNA-damage repair and unfolded protein response were all confirmed by RT-PCR to be implicated in the stress response to furfural[46]. As both compounds trigger complex stress response pathways, engineering efforts to increase tolerance to aldehyde inhibitors have instead focused on the inhibitor detoxification. For example, overexpression of the NADPH-dependant aldehyde dehydrogenase *ALD6*, which has a dual role of detoxifying furfural and 5-HMF, within *S. cerevisiae* led to increased cell growth and ethanol production[51]. The most successful attempt to increase furfural tolerance however was following a genome-wide RNAi screen which identified *SIZI1,* an E3 SUMO-protein ligase, as a knockdown or deletion target. Resultant phenotyping identified maximum specific growth rates of 0.26 h^{-1} for the *siz1*Δ mutant, versus just 0.15 h^{-1} for the wild type when grown within media containing 0.8 g/L furfural; increased tolerance was also observed to 5-HMF[52]. Of the non-conventional yeasts, *Pichia kudriavzevii* stands out for its tolerance to aldehyde inhibitors, still exhibiting growth in media containing concentrations up to 7 g/L furfural or 7 g/L 5-HMF[53]. Despite this exceptional tolerance, this species is unable

to metabolise xylose, cellobiose, arabinose and maltose making it difficult to be effective within a lignocellulosic process[42].

Tolerance to acidic inhibitors is mediated by maintaining cytoplasmic pH homeostasis through the action of proton pumps, as well as multidrug resistance transporters, to remove protons from the cell. *PDR5,* along with other members of the ATP-binding cassette family have has been implicated within *S. cerevisiae* in the expulsion of a wide range of toxic compounds, however *AZR1* has been found to contribute significantly to acetic acid tolerance with knock out mutants showing both reduced growth rates and increased lag times[54].

1.3.2. Fermentation inhibitor removal

At an industrial scale, a trade off scenario exists when pretreating a feedstock. Harsher methods which release the greatest amount of mono- and disaccharides generate the highest levels of inhibitors, milder pretreatment processes render lower inhibitor concentrations leave a high proportion of non-fermentable oligosaccharides. As harsher methods are usually preferred, a range of strategies exist to remove fermentation inhibitors before commencing cell growth. These include ion exchange resins[55], activated charcoal[56] and biological treatment through lignolytic enzymes[57]. Additional biological methods include the addition of tolerant species such as *Trichoderma reesei* which secrete cellulases and hemicellulases to assist with the more mild pretreatment processes[58]. These methods, however successful, add both cost, time and complication to the process and in order to compete on an economical level with less sustainable alternatives, these additional processing steps are effective but largely unfeasible.

1.4. Strain improvement and adaptive evolution.

1.4.1. Starting with the right strain.

Phenotypically improving *S. cerevisiae* to better suit the demands of a lignocellulosic bioprocess has often utilised a rational engineering approach. Here, the knowledge,

and subsequent modification of native biological systems can be performed to fine tune performance. Alternatively, strains of interest can be equipped with non-native genes or gene pathways allowing, for example, the metabolism of novel carbon sources. A stand out example of this within biotechnology is engineering the genes required for xylose metabolism within *S. cerevisiae* to improve ethanol yield from second generation feedstocks[59]. Although introducing and expressing novel pathways within a strain can itself be simple[60], a common problem encountered is that such modifications lead to 'off-site' effects on other tightly regulated biochemical networks present within engineered strains[61]. An example of this within the engineering of *S. cerevisiae* for xylose metabolism came from NAD(H) and NADP(H) cofactor imbalance, leading to poor substrate utilisation, biomass and productivity[62].

Due to the complexity and the range of the stress elicited from growth in the presence of fermentation inhibitors, rational design genetic engineering can be difficult to approach. For this reason, lignocellulosic bioprocesses have recently turned to non-conventional yeasts which can naturally tolerate high inhibitor concentrations and metabolise alternative carbon sources such as xylose, as potential platform organisms. One caveat however is that as non-conventional yeasts quite often do not have the appropriate genetic tools for strain improvement, it is important to only consider strains which possess the required phenotypes of a lignocellulosic process, for example being able to utilise all carbon sources. This, for example, would rule out the previously mentioned Pichia kudriavzevii[53] or Zygosaccharomyces *bailii,* which can tolerate up to 24 g/L acetic acid but cannot metabolise all of the sugars present within second generation feedstocks[63]. Where the necessary phenotypic range does exist however, strain development is often still required meaning that methods required need to look beyond rational engineering approaches.

1.4.2. Adaptive Laboratory Evolution.

Non-GM approaches are attractive in occasions where genetic modification routes are unavailable or unsuitable, and adaptive laboratory evolution (ALE) is such a tool for this. ALE is a method whereby microorganisms are continually cultured under a selective pressure in defined conditions for periods of time (weeks, to months or years), to allow for the selection of strains with desired phenotypes[64]. Whilst ALE is by no means a novel method for phenotypic improvement, recent advancements in sequencing technologies have meant that advantageous phenotypes can now be queried genotypically.

ALE methodologies typically fall within two categories: sequential batch culturing or continuous chemostat culturing, with either method having advantages or disadvantages over the other (Figure 7). Sequential batch culturing requires the serial transfer of cells from one batch to another at a pre-determined regularity, for instance when a certain OD is reached, a given number of generations, a defined length of time, or at a particular stage in growth. The design of such experiment, given its hands-on nature, is often a trade-off between an optimal transfer time and the 'user friendliness' of the set up. Key fitness parameters monitored within batch ALE experiments include length of lag phase, growth rate, and the density after a given time of growth. Advantages of the sequential set up is that batches are run in parallel, allowing the opportunity for common, or different, mutations to appear across different lineages, or additionally means that any cell lines which do not survive can be replaced by successful ones[64]. Disadvantages are in the lack of control of culturing conditions beyond what can be defined at point of transfer to a new batch. Chemostat evolution strategies in comparison aim to maintain a constant growth rate through media addition and culture removal. Here, regulation of pH, dissolved oxygen and nutrient concentration can be maintained to ensure that cells are kept in exponential phase, and in doing so providing the ideal backdrop for evolution. As a set-up such as this requires specialist equipment, it is not possible in all instances and for the same reason cannot be used to generate independent cell lines to compare evolutionary routes to strain improvement[64].

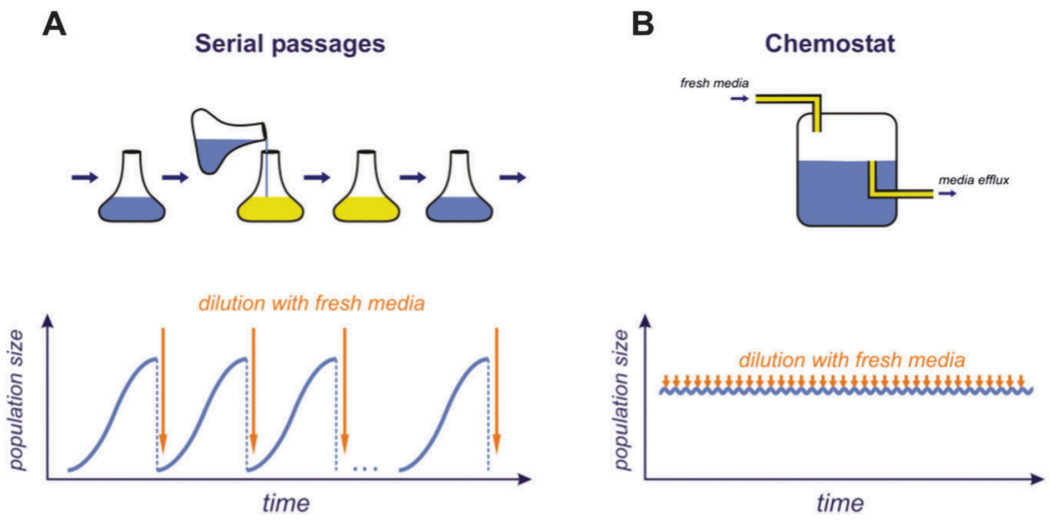


Figure 7. Comparison between sequential batch culturing and chemostat culturing for ALE. A – Sequential batch culturing is based on the serial transfer of cells at a defined period of time into fresh media. This approach has the advantage of being able to evolve several cell lines in parallel, allowing the comparison of acquired mutations. The batch approach however has a reduced capacity to control culturing conditions outside of what can be defined within the media at the start of a batch. B – Chemostat evolution uses an advanced set-up to maintain cell population size and growth rate by adding/removing media at a constant or adaptive rate. Conditions such as pH and dissolved oxygen can also be better controlled in this system, however due to the comparative complexity of chemostat evolution, this process is often not performed in parallel reducing the capacity for comparative genomics. Figure from Mozhayskiy[65].

1.4.3. Whole genome sequencing and the mechanisms of evolution.

Inexpensive whole genome sequencing and development of analysis software has transformed the impact of adaptive evolution experiments. Now, the genetic basis or causality underpinning improved phenotypes can be determined, making this a powerful tool within reverse engineering strategies[66]. In comparison to rational engineering which begins with an a priori strategy, reverse engineering begins with a strain improved for a specific phenotype, and attempts to apply the genetic bases for the improvements into a wildtype or lab strain. When approaching the improvement of a particular phenotype, rational engineering, as previously discussed, can lead to undesirable off site effects such as co-factor imbalance within strains of *S. cerevisiae*, or not follow the expected outcome[62]. In contrast, within evolved strains it is possible to have 'fixed' or worked around issues such as co factor imbalance, or undergone phenotypic improvement through routes unlikely to be

predicted through rational engineering[67]. An alternative is to combine both methods and use adaptive evolution to 'fine tune' rational engineering modifications. One such example took an industrial strain of *S. cerevisiae* engineered with the genes required for xylose metabolism, performed mutagenesis and genome shuffling before finally subjecting strains to sequential batch transfers in xylose containing media. Phenotypic evaluation after evolution identified increased xylose isomerase activity, possibly coming from multiple amplifications of the inserted *XylA* gene.

Mutations found within ALE studies which confer advantageous phenotypes include SNPs, small insertions and deletions, transposon movement or amplification, deletions or amplifications of larger genomic regions and changes in ploidy or emergence of aneuploidy[68,69]. These mutations culminate to cause increased/decreased gene expression, loss/improvement of gene function, novel gene function, or to change gene regulation. The mutations, and their frequency, are influenced by a species/strains specific variability in mutation rate. This in itself can be influenced by where in the genome the mutation has occurred based on variable DNA mismatch repair efficiencies at different regions[70]. In addition, genes with increased expression correlated with an elevated mutation rate[71]; a factor particularly relevant within ALE. For an advantageous mutation to become fixed within a population, it has to overcome the background population. For instance, a beneficial variant will have to overcome genetic drift; the change in frequency of an allele within a population due to random sampling[72]. This effect is particularly relevant within batch transfer ALE studies as cells are randomly carried to the next batch regardless of fitness or genotype. Secondly, ALE studies are likely to provide opportunities for clonal interference; a scenario where by two cell lineages possessing beneficial mutations interfere with the others dominance within the population, slowing down eithers further progression[73].

Though many forms of mutations are possible, ones which elicit an effect on gene dosage through copy number variants (CNVs) are among the most common within ALE studies. CNVs can range from the segmental addition or deletion of 50 base pairs

to entire chromosomes (Figure 8) [74]. Amplification of gene copy number can result in either functional divergence of the new copy or simply increase the dosage through heightened expression. Increased gene dosage is a common outcome in evolution experiments involving nutrient limitation, such as the specific amplification of the high-affinity sulphur transporter *SLU1* in evolution in sulphur limited media, or nitrogen transporters *PUT4*, *DUR3* and *DAL4* in nitrogen limited media[75]. CNVs where genes or sections of chromosomes are lost results in loss of heterozygosity events; reducing gene dosage or even resulting in further genome instability if genes lost are key to the maintenance of correct DNA copy number[69]. Most CNVs occur through errors in homologous recombination or non-homologous recombination during DNA repair. Homologous recombination is commonly required to repair mechanical stress events such as collapsed replication forks or strand breaks, and errors by this mechanism can lead to both duplications and deletions[74]. CNVs can also arise from non-allelic homologues recombination, which can occur when enough sequence similarity is present between sections of DNA as a result of previous gene duplication[76]. CNVs arising from non-homologous recombination occur during non-homologous end joining or microhomology-mediated end joining, where a DNA break occurs and can lead to the introduction of free DNA (transposable elements or mitochondrial DNA) or the loss of some genetic material[74]. Due to their respective sizes, CNVs arising from homologous recombination are more likely to implicate a number of genes. In contrast, fork stalling and template switch is a mechanism capable of generating CNVs within larger sections of the genome. Here, when the replication fork stalls due to stress, the 3' end of the DNA strand can change its template to a single stranded template of a nearby replication fork. This process can lead to regions of 20 kb being incorrectly amplified[74,77]. Furthermore, it has been shown that the consistent environmental conditions within ALE experiments promote the formation of CNVs within highly transcribed regions, caused by collisions between transcription and replication machinery[78]. Depending on which strand transcription or replication is occurring, collisions can be head-on or co-directional, with the former inhibiting fork movement greater than the latter and therefore causing more genomic structural changes[79].

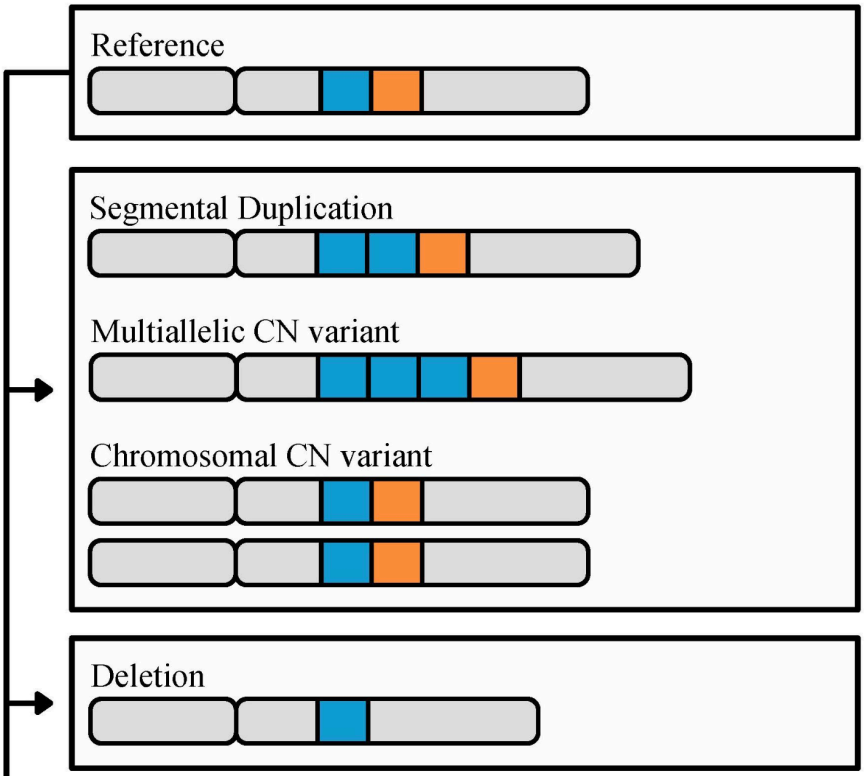


Figure 8. Different types of CNV. Relative to the reference chromosome containing two reference loci, different CNV can be achieved. Small scale segmental duplications are likely to create an additional copy of a locus, though the same process can yield a multiallelic variants to further increase copy number. Larger scale mutations can include the duplication of entire chromosomes, whilst deletion of a locus can cause a reduction in gene CN. Figure by Steenwyk[74].

There are many examples of successful ALE studies within literature, including strains of *S. cerevisiae* which have been adapted to have increased tolerance to ethanol [80], acetic acid[81] and the cocktail of inhibitors presented within hydrolysed lignocellulosic feedstocks[82]. Temperature tolerance can also be improved through ALE, and in one study an early stop codon within *ERG3* has been shown to cause a 1.5-2 fold increase in growth rate at 40 °C within a *S. cerevisiae* strain [67]. Finally, ALE has been used to improve aspects of carbon metabolism, for example SNPs to the *GUT1* and *UBR1* genes have been identified to confer increased glycerol metabolism, and SNPs within hexose transporters *GUT2*, *HXT5* and *HXT7* have generated glucose-insensitive xylose transporters[83,84]. ALE has also been successfully applied in non-*Saccharomyces* yeasts. The oleaginous bacterium *Rhodococcus opacus* was adapted to a fermentation inhibitor cocktail containing

lignin, 4-hydroxybenzaldehyde and syringaldehyde, which resulting evolved strains exhibiting a markedly shortened lag times[85]. Duplication of *XYL1* and *XYL2* has been identified as causative within strains of *Yarrowia lipolytica* with improved biomass and lipid production using xylose as the sole carbon source[86]. Finally, a strain of the oleaginous yeast *Rhodosporidium toruloides* evolved within sugarcane bagasse hemicellulosic hydrolysate displayed increased lipid production and reduced lag times when grown within this feedstock than the progenitor[87].

1.4.4. Evolutionary Trade-offs

Though clearly an effective route to industrial strain development, ALE, and the genetic changes underpinning adaptations, can lead to reduced fitness when strains are cultured outside of these conditions. For instance, within the *EGR3* mutated high-temperature tolerant *S. cerevisiae* strain discussed above, the evolutionary process also rendered the resulting strain respiratory deficient, with reduced growth rate at 30 °C[88].

Due to the environmental and therefore transcriptional consistency elicited by ALE studies, deleterious SNPs or CNVs to parts of the genome can be masked and allowed to accumulate with little deleterious effect on the strain. This effect may be particularly significant in ALE studies to improve utilisation of a non-preferred carbon sources, where there are comparatively few candidate genes to which beneficial mutations or CNVs can occur. In *S. cerevisiae* strains which have undergone ALE for improved glycerol utilisation, were found with reduced osmotolerance or hampered ethanol utilisation[89]. A similar effect has been shown to occur when *S. cerevisiae* strains were adapted to glucose limitation, with resulting strains unable to compete with the progenitor when carbon concentrations were no longer limited[90].

Clearly therefore, the successes of ALE is dependent on the wider requirements of the strain. For example, successfully improving tolerance to acetic acid coming at the expense of reduced utilisation of the carbon sources within lignocellulosic feedstocks does not make for a better overall strain. Recently however, dynamic ALE strategies

have been successfully applied to alleviate such potential trade-offs. Here, the growth of evolving strains was alternated between glucose, xylose and arabinose in single sugar batch format to improve the overall fermentation of mixed sugar cultures[91]. In this context, should a strain generate mutations beneficial for xylose metabolism at the expense of glucose, then it would not continue within the progressing cell population due to this dynamic strategy.

1.5. High throughput lipid quantification.

An accurate and reliable lipid quantification method is essential when investigating oleaginous microorganisms. The Bligh and Dyer method, though developed in 1959 for the extraction of lipids from frozen fish samples, remains the standard for lipid extraction and quantification within the field, though now widely modified for microbial cells[92]. Adapted methods still comprise of a two-step process; cells initially undergoing some form of lysis (mechanically, chemically or enzymatically) followed by extraction of the released neutral lipids using a nonpolar organic solvent such as chloroform. The extracted cellular debris remains within the aqueous layer. Though simple in concept, in practice this method is low-throughput, requires the use of organic solvent and has been shown to consistently underestimate lipid yield as the dry weight percentages increases[93,94]. For these reasons, the use of traditional lipid extractions is unsuitable for high throughput studies, or in studies where not enough biomass can be recovered for an accurate extraction.

1.5.1. Nile red fluorescence

Better suited to high throughput experiments are lipid specific, fluorescence techniques using lipophilic dyes. For this, Nile red is the most commonly used, and has been used to estimate lipid yield within yeast, fungi, microalgae and ciliates[95–100]. Nile red is highly fluorescent within hydrophobic organic solvents, but quenched almost absolutely within aqueous solutions. Nile red also fluoresces brightly within, and is highly specific to, the neutral lipids found within the lipid droplets, avoiding others such as membrane phospholipids making it perfectly suited as an oleaginous lysochrome[99]. Despite this, the emitted fluorescence can vary depending on the

fatty acid, with unsaturated fatty acids giving a stronger measurement than saturated under the same conditions[101]. A mechanistic study identified a two-step incorporation of Nile red into exposed cells, with a shifting a fluorescence peak shift indicating an initial immediate entry into the plasma membrane (625 nm) followed by a slower transition of the stain into lipid droplets (580 nm) (Figure 9) [102].

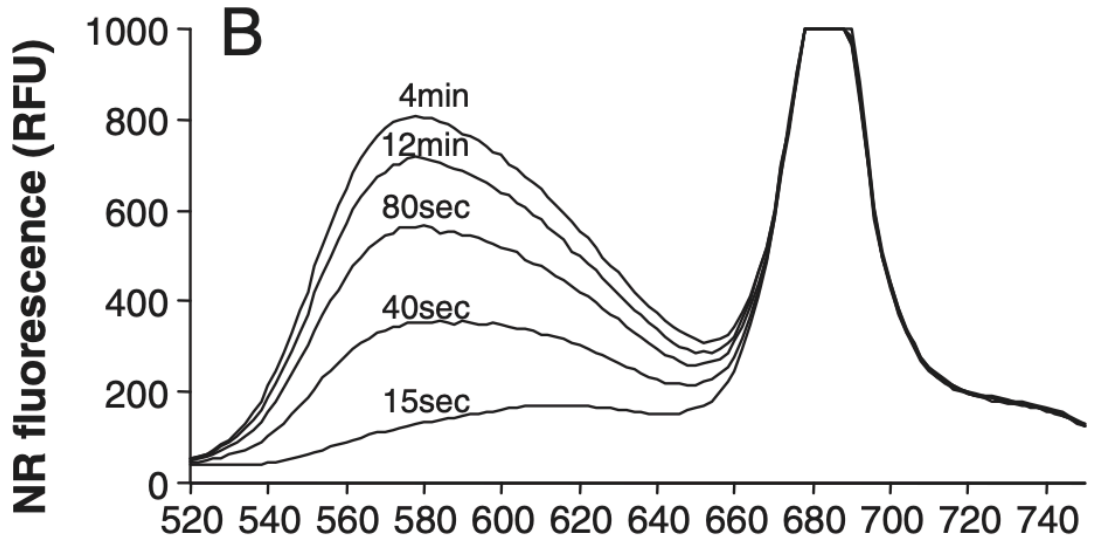


Figure 9. Time-course analysis of Nile red staining in microalgae. Upon exposure of microalgae cells to Nile red, an initial peak occurs at 625 nm as the stain enters the plasma membrane, followed by the transition of the stain into lipid droplets, identified by a peak at 580 nm. Figure from Pick[102].

Though unable to provide the quantitative weight percent result of a traditional lipid extraction, fluorescent dyes can be used for several types of analysis including fluorescence microscopy, spectrofluorometric high throughput screening and flow cytometry[97]. Methods have been developed enabling semi-quantitative lipid production data via comparison of oleaginous lipid straining with a neutral lipid standard curve, usually with the view to apply this within a microplate format. These studies generally involve the use trioelin as the lipid standard; generating 'trioelin equivalent' results to quantitatively compare samples[103]. Milk fat has also been used as an alternative, as it is reported that trioelin forms lipid droplets at higher concentrations, leading to impaired measurements[104]. Flow cytometry together with Nile red fluorescence staining can also be used to provide an estimation of lipid production, and has been shown to correlate well with traditional lipid extraction

data for yeast and microalgae[105–107]. In addition, flow cytometry has the advantage of being able to identify intensity per cell, and in conjunction with Flow Particle Image Analysis as a method to determine individual cell size, can provide valuable information of population dynamics as cultures progress (Figure 10) [105].

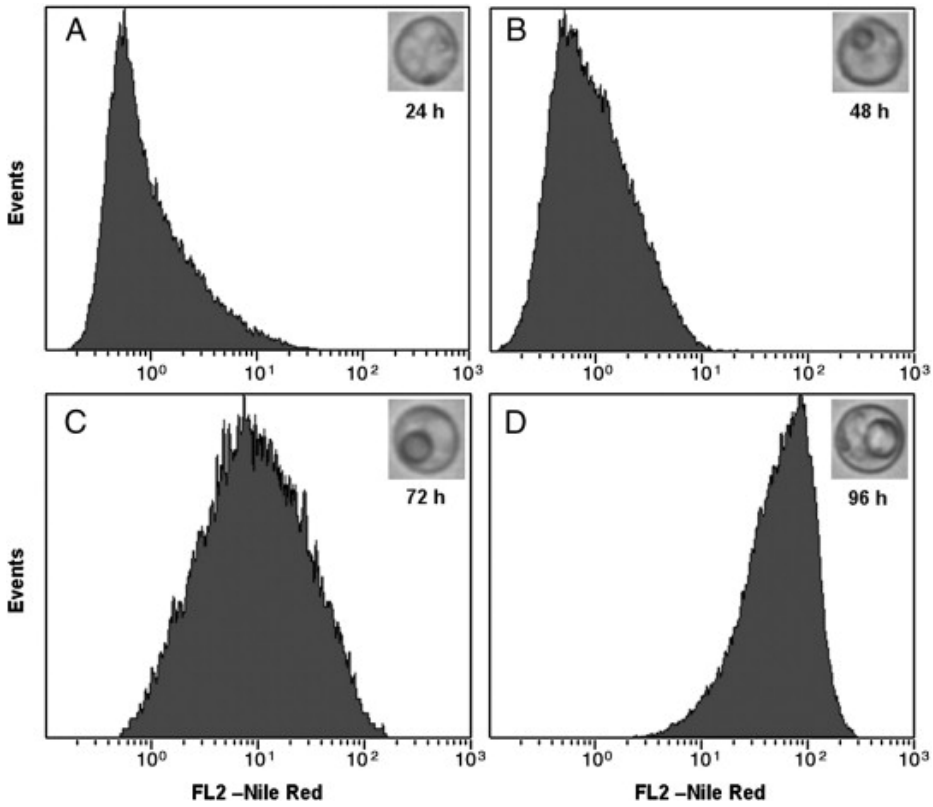


Figure 10. Distribution of per-cell lipid droplet fluorescence as determined by flow cytometry. Data shows time-course sampling of an oleaginous yeast culture as lipid accumulation occurs. Data also gives an indication of population dynamics with respect to per-cell lipid accumulation. Figure from Raschke[105].

1.5.2. Inconsistencies of Nile red methods.

Though widely used within the study of oleaginous microorganisms, the are several limitations to the effective and consistent use of fluorescent staining for estimation of lipid production. These factors fall predominantly within two main categories: variation of dye diffusion into cells between different strains or species, and fluorescence quenching.

Factors influencing effective permeation of dye into lipid droplets include cell size, lipid droplet size/number, dye interactions with proteins or other cytoplasm components and most of all, cell wall thickness[97]. The latter has been identified as particularly significant as both lipid accumulation and increasing cell wall thickness generally take place within the stationary phase[108]. Indeed, varying cell wall thicknesses during different growth phases of microalgae has led to poor correlations between Nile red fluorescence and traditional lipid extractions[96]. A similar conclusion has been drawn within yeast, where it was determined that different species took varying amounts of time to achieve maximum fluorescence. Here, introducing a 20 minute kinetic reading rather than a single time point to the staining method was determined to allow for this inherent strain-strain variation[99]. To increase stain permeation, DMSO (dimethysulfoxide) is frequently added due to its interaction with the cell membrane; either by creating pores within the lipid bilayer or modifying membrane fluidity[109].

Fluorescence quenching, characterised by a peak-florescence measurement being reached followed by a gradual decline in overall intensity, has been reported to occur in yeasts[98,102]. In one study, it was shown that the peak fluorescence reached after 5 minutes was rapidly lost, with just 10% of the maximum remaining after 20 minutes post Nile red exposure (Figure 11A)[102]. Further analysis within the same study revealed the speed of fluorescence decay was highly varied amongst different oleaginous species, with one strain retaining >50% of its maximum even after 120 minutes (Figure 11B)[102].

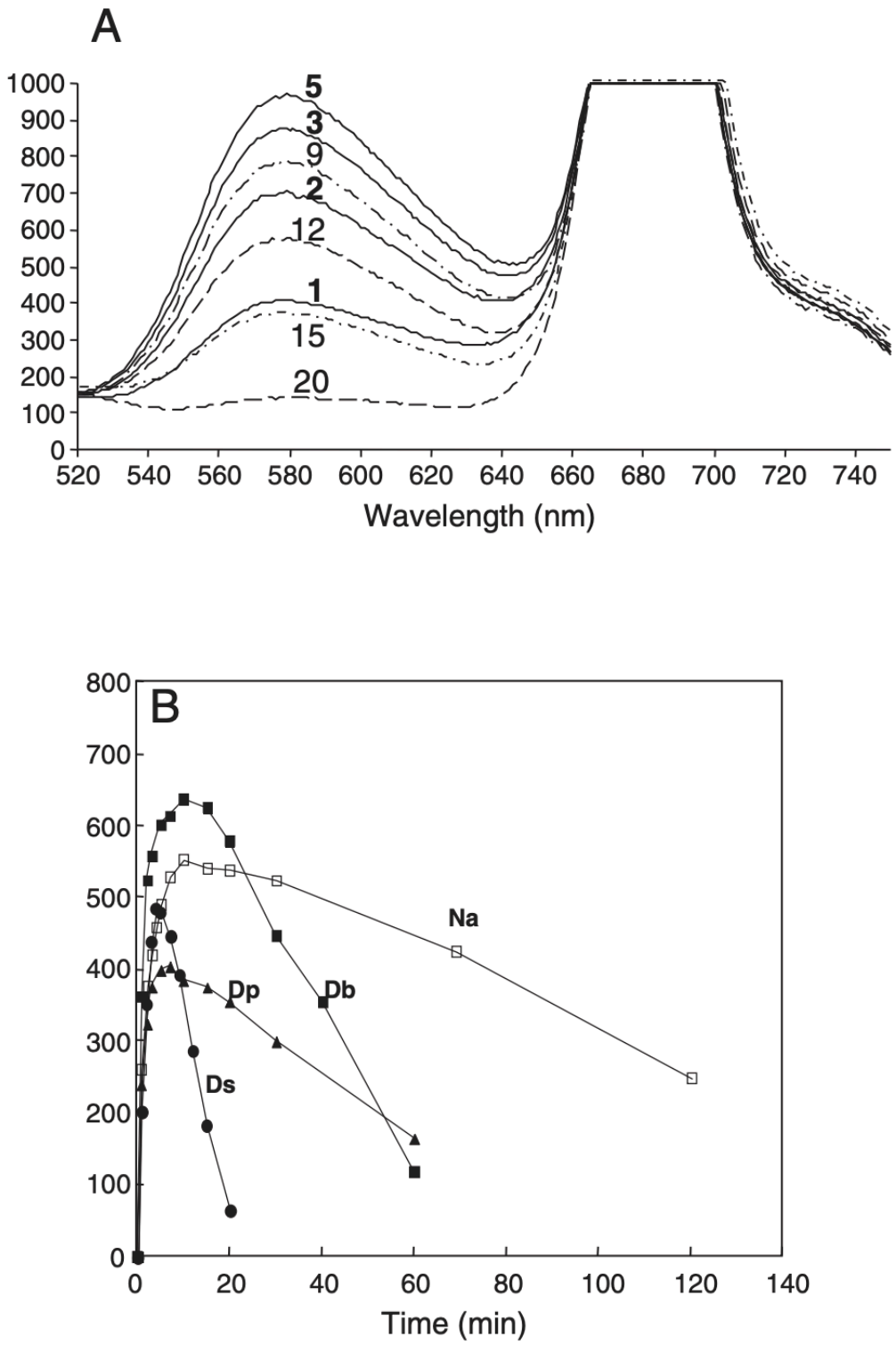


Figure 11. Quenching of Nile red fluorescence overtime and strain dependant variability in quenching. A: After reaching peak fluorescence after 5 minutes, fluorescence gradually drops over preceding measurements, to reach baseline after just 20 minutes. B: Fluorescence quenching overtime in four different stains of microalgae. Na - *Nanochloris atomus*, Dp - *Dunaliella parva*, Db - *Dunaliella bardawil*, Ds - *Dunaliella salina*. Figures from Pick[102]

Though the possible method inconsistencies deriving from fluorescence quenching are obvious, the mechanism causing quenching is not known for certain. It has been shown that the degree of quenching correlates with increasing Nile red concentration, with quenching almost removed when the ratio of Nile red to cells is low[102]. In addition, glycerol was shown to play a key role in the concentration-dependant quenching of Nile red, where it could act to both improve stain solubility and also quench fluorescence. The conclusion drawn within this study stated that when Nile red concentrations are low, stain arrangement with respect to the lipid droplet masks them from attack from hydrophilic (such as glycerol) quenchers, with the reverse true when stain molecules stack within droplets when dye concentrations are high (Figure 12)[102]. As intracellular glycerol levels depend on environmental cues such as the need for osmoregulation or redox balancing, degrees of quenching may be difficult to predict when using a defined staining protocol[110]. The flux of glycerol within oleaginous strains could conceivably be even more varied due to the role glycerol plays within TAG biosynthesis, complicating the quenching effect further.

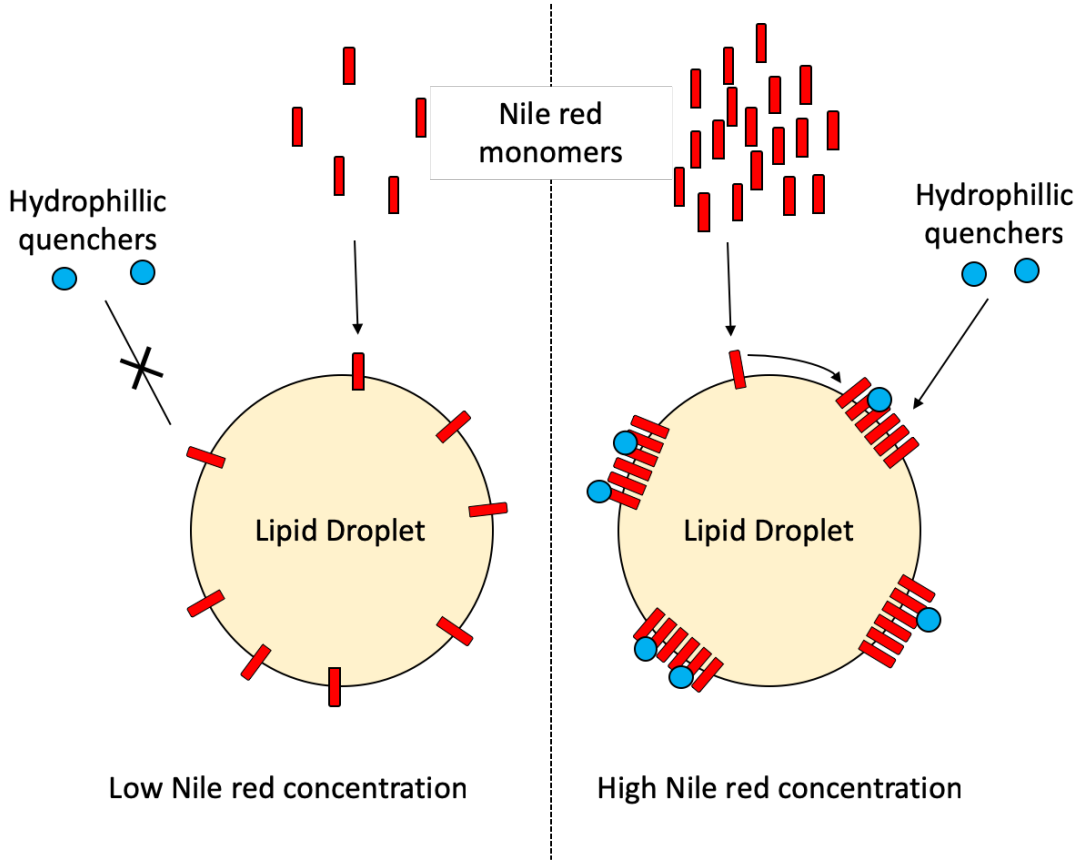


Figure 12. Predicted model of Nile red quenching. When concentrations of Nile red are high, Nile red monomers arrange into exposed oligomers on the lipid droplet surface. In this confirmation, Nile red oligomers are open to quenching and a decrease in fluorescence is seen over time. When Nile red concentrations are low, it is found that quenching does not occur. Glycerol has been identified as one of the main causes of hydrophobic quenching. Figure adapted from Pick[102].

1.5.3. Non-Fluorescence based lipid estimation.

Due to the inconsistencies of fluorescence-based methods, alternative methods of lipid quantification have been developed. Further to simply providing an estimation of production, these methods often seek to provide information in addition to what can be gathered through Nile red.

Fourier Transform Infrared Spectroscopy (FTIR) has been shown to generate quantitative lipid data as well as identification of different lipid classes[111,112]. In the first instance, tracking IR absorbance of lipid indicative CH_2/CH_3 (3000-2800 cm^{-1}) and C=O (1740 cm^{-1}) bonds over time has been able to accurately track lipid accumulation (Figure 13). Further analysing absorbance data can also provide information on specific lipid properties, for instance emergence of =CH bond absorption gives an indication of unsaturated fatty acid formation. This work also showed that thickening of the cell wall through the increase of beta-glucans can also be observed at the region between 1250 and 1000, which can too correlate with lipid accumulation[112]. Raman spectroscopy, though providing less lipid specific analysis versus FTIR, has also been used as an extraction free method of lipid quantification in microalgae, as well as providing additional chemical fingerprinting of chlorophyll and carotenoid abundance[113].

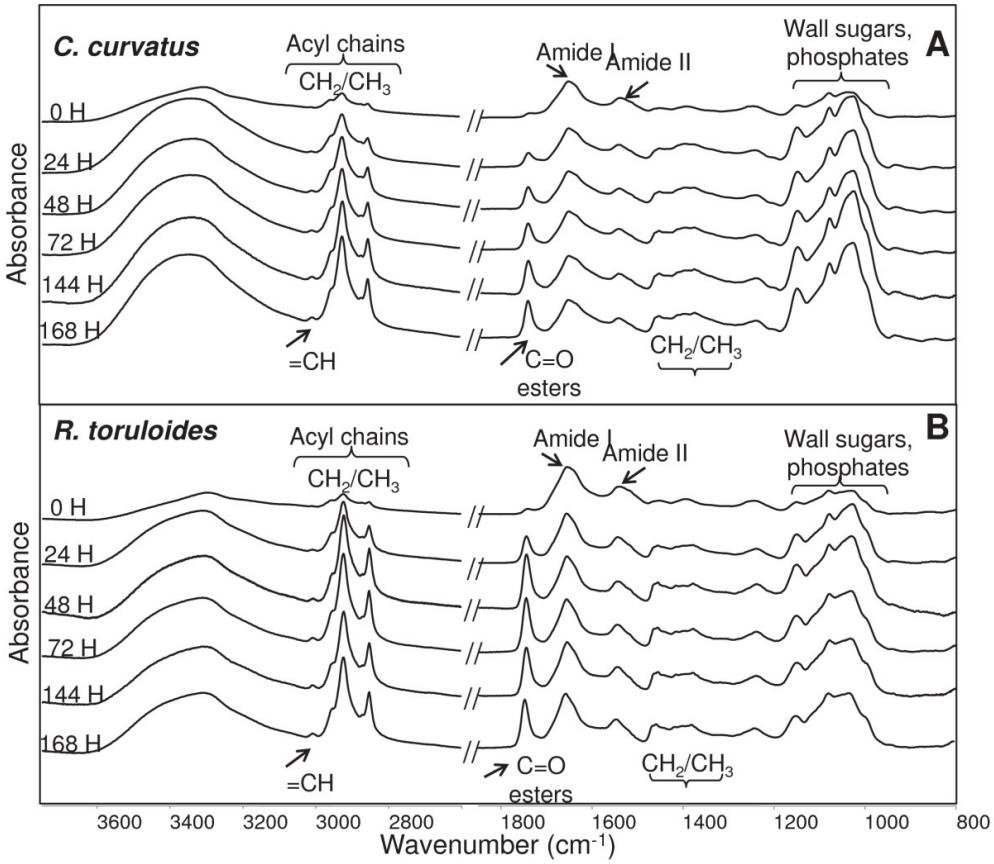


Figure 13. Fourier Transform Infrared Spectroscopy of lipid droplet accumulation. IR spectra for increased absorbance of CH_2/CH_3 (3000-2800) and C = O (1740), two lipid indicative adsorption bands, over time course samples is able to accurately depict lipid accumulation. Figure from Ami[112].

Finally, high content image analysis has recently been applied to oleaginous yeast to quantitatively estimate lipid accumulation within a population and provide information of single cell lipid dynamics, all in an automated manner. Using Nile red and fluorescein staining together, lipid droplet volume, estimated lipid weight per cell, lipid droplet productivity and cellular distribution of lipid droplets were able to be determined (Figure 14)[93].

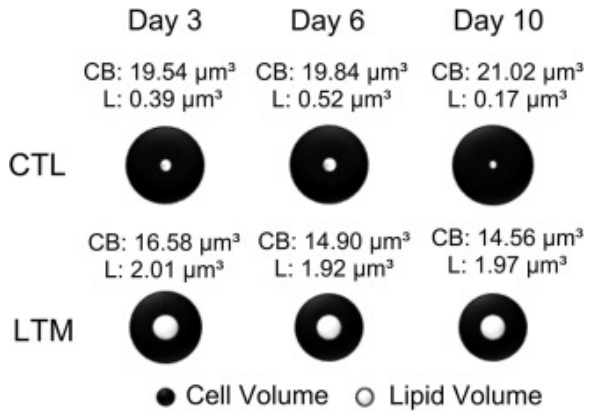


Figure 14. High content image analysis of a oleaginous yeast. Representations of two different cell types, a control (CTL) and a lipid producing strain (LTM). Image analysis software can determine average lipid droplet area and average cell size. Figure from Capus[93]

1.6. *Metschnikowia pulcherrima*: a potential platform organism.

Within this PhD study, the oleaginous yeast *Metschnikowia pulcherrima* will be investigated as a potential platform organism[15]. Widely found on flowers, fruits and plants, this xylose fermenting yeast has previously gathered commercial interest due to its natural biocontrol properties stemming from its inherent antimicrobial activity[114]. One study investigating the selectivity of *M. pulcherrima's* antimicrobial activity showed that when screened against 114 strains representing 32 different species such as *Saccharomyces, Pichia, Candida* and *Kluyveromyces,* one strain of *M. pulcherrima* displayed inhibitory effects to all except the S*accharomyces* strains. Strains of *M. pulcherrima* are also shown to be effective in inhibiting clinical isolates of *Candida* spp*.,* including *C. albicans* and *C. tropicalis*[115]*. M. pulcherrima's* antimicrobial activity is thought to derive from its production of pulcherrimin; an irreversible chelator of environmental iron. Removal of iron from growth conditions means that competing organisms display inhibited growth[114,116]. Antimicrobial activity could also derive from naturally high production of alcohols such as 2-phenylethanol. This organism has previously been investigated as a platform organism for the commercial production of this compound, and titres of 1500 mg L^{-1} have been achieved within an optimised 2 L bioreactor system[117]. This titre

represents the highest reported de novo titre to date, indicating strong commercial? potential for this highly valued flavour and aroma compound.

M. pulcherrima's ability to use a broad range of metabolisable carbon sources, particularly of the main three saccharides found within lignocellulose; xylose, glucose and cellobiose, make it an ideal yeast for exploitation within biotechnology[118]. Furthermore, the metabolism of lignocellulose derived oligosaccharides, including glucans and xyloglucans up to a chain length of seven, by this organism has been demonstrated when grown on hydrolysed lignocellulosic material[119]. Wonder if*M. pulcherrima* has also demonstrated tolerance to high concentrations of the fermentation inhibitors furfural, acetic acid and 5-HMF (1, 2.5 and 2 g/L respectively). In fact, the inhibitor tolerance of this organism compares favourably to strains such as *Yarrowia lipolytica*, *Rhodosporidium toruloides* and *Lipomyces starkeyi* and in total outperformed 40 others when grown in these inhibitors[118]. Furthermore, *M. pulcherrima* has shown to produce lipid with a composition similar to that of palm oil when grown on waste feedstocks within non-sterile conditions, and has been shown within the same study to produce 40% lipid by dry weight within optimised conditions[15]. *M. pulcherrima* also has an additional benefit in that there is no intellectual property covering its commercial use as an oleaginous organism, except for a that filed by this research group, meaning there is suitable freedom to operate. These factors, in combination with how this species is regarded as a safe organism, give it great potential for investigation.

1.7. Project Aims.

Work within this thesis is split into two work flows.

1. Development of a high-throughput lipid quantification method.

To provide absolute weight percent quantification of lipid accumulation, there is currently no reliable alternative to traditional lipid extraction. Though this method will be used throughout, an alternative lipid estimation method will be developed for use within this organism. A fluorescence-based screening method will be optimised,

however given the discussed inconsistencies of such stains, non-fluorescence avenues will also be explored. The developed method will seek to:

- Be appropriate for high-throughput sampling,
- Allow for reliable comparisons to be made between separate experiments,
- Be robust to environmental factors or culturing time,
- Provide additional information to accompany estimation of lipid accumulation.

2. Adapted laboratory evolution to improve industrially relevant phenotypes.

At present, the genetic toolkit for *M. pulcherrima* is being developed within the group. For this reason, work within this thesis will utilise ALE strategies for strain improvement.

For an economically viable process to be successful, low cost lignocellulosic feedstocks will be used as the basis of the bioprocess. Although the purpose of the project is to utilise *M. pulcherrima* as a platform organism for the production of lipids, there are various phenotypic traits which require improvement should this organism be suited to growth on lignocellulosic feedstocks. ALE will be used within this study to improve two phenotypes, tolerance to fermentation inhibitors and xylose utilisation.

A) Improving tolerance to fermentation inhibitors

Two strategies of ALE to improve tolerance to fermentation inhibitors will be performed, starting with the same progenitor strain; adaptation to a cocktail of fermentation inhibitors (formic and acetic acid, furfural and 5-HMF) or adaptation to a single inhibitor (acetic acid). Resulting strains will be phenotypically characterised and whole genome sequenced to attempt to find causative mutations driving

adaptation. By adopting different strategies, the following questions will be investigated:

- Will genetic routes to adaptation be specific or generic? I.e. will the single inhibitor adapted strains have mutations specific to acetic acid and the inhibitor cocktail adapted strains more general adaptations?
- Will different phenotypic trade-offs occur between the two strategies.
- Will the single inhibitor adapted strains have improved growth when cultured in an inhibitor cocktail?

B) Improving xylose metabolism

Xylose represents a significant proportion of the available carbon source within many secondary feedstocks and its effective metabolism and conversion into product or biomass is crucial to an industrial bioprocess. *M. pulcherrima* has been shown to metabolise xylose, however conversion into biomass is well below that of other carbon sources such as glucose, glycerol and cellobiose[15,118]. Xylose consumption has also been shown to be an extremely long process, taking 12 days to be fully metabolised within a mixed sugar culture[15]. Here, ALE will be performed using xylose as the sole carbon source. The resulting strains will be phenotypically characterised and subject to whole genome sequencing to uncover underlying genetic mechanisms of adaptation.

By performing these two ALE strategies, the general adaptive mechanisms of *M. pulcherima* will be discussed. Though a similar experimental design will be followed, the selective pressure driving adaptation to fermentation inhibitors is physiologically different to that presented by adaptation to xylose. Therefore, whilst the ALE studies in this thesis will be primarily to improve the industrial phenotypes of *M. pulcherrima,* it will also enable the study of the evolutionary mechanisms driving adaptation when posed different types of selective pressure.

2. Comparison of Nile red and cell size analysis for high throughput lipid estimation within oleaginous yeast.

2.1. Commentary

Lipid extractions of oleaginous yeast is the go-to method to quantify lipid accumulation. Although this method is accurate and reliable, it is not conducive to high throughput experimentation. As work with *M. pulcherrima* prior to that within this thesis was mainly performed at larger scales, a high throughput method was not required, meaning no work had been performed to develop one. The scope and quantity of experiments planned within this thesis would however require a high throughput lipid estimation method.

The work presented in this chapter details both the development of a novel, fluorescence-free method of high throughput lipid estimation as well as the optimisation of a Nile red fluorescence-based method. Though presented as a complete method here, a lot of experimentation was conducted into the use of Nile red outside of this publication before reaching that point. After realising the almost unavoidable inconsistencies with the Nile red method, it became clear that an alternative was required. Initial work into an alternative considered a combination of cell imaging and fluorescence, with the aim of using images to measure the area of lipid droplets and to use this as a proxy for lipid quantification. The rationale here was that even though the inherent staining issues meant that cells with large lipid droplets were giving a far lower fluorescence measurement than they should, there was still staining occurring. This could mean that however faint the fluorescence, an area could still be measured. Unfortunately, the issue of variable fluorescence quenching meant that although this could work for some samples, fluorescence was often quickly lost in cells with large droplets.

During work to develop a lipid droplet area method of quantification, it was realised that overall cell size alone was potentially giving an accurate depiction of lipid droplet size. After some successful initial experiments into this potential method, the following piece of work was written which presents this alternative fluorescence free

method, validates it against traditional lipid extractions, and compares it with an optimised Nile red method.

Comparison of Nile red and cell size analysis for high throughput lipid estimation within oleaginous yeast.

Authors: Robert H. Hicks[a], Christopher J. Chuck[b], Roderick J Scott[c] , David J. Leak[c] , Daniel A. Henk[c].

[a]Corresponding author. Centre for Doctoral Training in Sustainable Chemical Technologies, Department of Biology and Biochemistry, University of Bath, Bath, BA2 7AY, United Kingdom. Email: R.H.Hicks@bath.ac.uk

[b] Department of Chemical Engineering, University of Bath, Bath, BA2 7AY, United Kingdom

[c] Department of Biology and Biochemistry, University of Bath, Bath, BA2 7AY, United Kingdom

Running title: Cell size analysis for high throughput lipid estimation

Key Words: Oleaginous Yeast, Nile Red, Microbial Oil, High-throughput Screening, Microscopy.

Abbreviations:

DMSO - Dimethyl sulfoxide

CSA – Cell Size Analysis

Abstract.

With growing interest in oleaginous yeast as producers of future fuels and bulk chemicals, a robust, high-throughput method for estimating lipid production is required. Although the lipophilic dye Nile red is frequently used to assay large samples of yeast and microalgae, inconsistent stain permeability between species and strains limits its effectiveness for some microorganisms. In this study, the oleaginous yeast *Metschnikowia pulcherrima* was used to develop a fluorescence-free, cell size based image analysis method for estimating lipid production, which was then compared with an optimised Nile red method across several experimental scenarios. Cell size analysis outperformed Nile red in all scenarios, correlating well with lipid extraction data when screening multiple strains, screening a subset of strains grown in different conditions and tracking the lipid accumulation of a culture over time. Stain permeability was shown to vary significantly among the strains trialled, with lipid droplet size and cell wall thickness having a deleterious effect in the permeability of high lipid accumulating cells. Cell size analysis could also allow culture population dynamics to be monitored, providing key process information of cell size distribution in response to changing media compositions.

Practical applications:

Nile red is currently the go-to method for high throughput lipid screening, however staining inconsistencies in some organisms caused by varying cell morphology makes it challenging to optimise a robust protocol. Though fluorescence free methods exist (Raman spectroscopy, FTIR, GCMS) the need for extensive sample preparation and specialist equipment restricts their widespread adoption. The cell size analysis method presented here offers an accurate, robust and cheap alternative for the study of microorganisms where fluorescence based avenues are not feasible. Furthermore, the population dynamics collected during cell size analysis can easily be applied to bioreactor style processing, where tracking size distributions can provide real time information of culture status. This additional information is valuable even if fluorescence screening is a possibility.

2.2. Introduction

Microbial lipids produced by oleaginous organisms represent a sustainable alternative to fossil derived fuels and chemicals. Furthermore, due to their comparable lipid profile, oil produced by oleaginous yeasts represents a viable replacement for plant oils, and because of the increasing biotechnological interest in this area, there is a growing need for fast and reliable screening methods of lipid production [1]. The widely used Bligh and Dyer method of lipid extraction and quantification is not suitable for high throughput analysis, lipid screening methods have instead largely focused on fluorescent staining. Amongst oleaginous yeasts, Nile red is the most commonly used lysochrome due to its high specificity for neutral lipid, allowing spectrofluorometric assays to be developed [2]. Though unable to provide a quantitative mass ratio, methods developed to include a trioelin standard curve have allowed for quantitative analysis and comparison of assayed samples [3].

Despite its widespread use, inefficient dye permeation and the need for method optimisation influence the reliability and efficacy of fluorescence staining protocols. Permeation issues reported within oleaginous yeasts and microalgae are likely to be caused by differential cell size, lipid droplet size/number, protein-dye interactions and most of all, cell wall thickness [2,4]. The influence of cell wall thickness on dye permeation is particularly significant here, as both lipid accumulation and cell wall thickening are known to occur during the stationary phase of yeast [5,6]. Indeed, poor correlations with lipid extractions are reported within macroalgae when staining samples at different growth stages [7]. Furthermore, during the development for an improved Nile red fluorescence assay, Sitepu reported a high variability in the time required for different species to reach maximum fluorescence emission, in addition to inconsistencies as to whether the addition DMSO as a stain carrier yielded an increased measurement [2]. A similar observation was found within microalgae, where Balduyck reported method optimisation as species dependent [8].

Non-spectrofluorometric extraction-free lipid assays using flow cytometry or image analysis have been published, however as they also require fluorescence staining it follows that inconsistent dye permeation can still limit their reliability with some

organisms [9,10]. Fluorescence free methods on the other hand often require specialist equipment such as Fourier Transform Infrared spectroscopy, Raman spectroscopy or time-domain NMR, as well as extensive sample preparation [11–13].

In this study, a fluorescence-free, cell size based image analysis method for the estimation of lipid production was developed and validated against traditional gravimetric analysis. The method was also assessed against a published Nile red assay further optimised for this study.

2.3. Methods

2.3.1. Chemicals

Unless otherwise stated, chemicals were sourced from Sigma Aldrich and used without further purification.

2.3.2. Strains, strain maintenance and media

Six *Metschnikowia pulcherrima* strains representing different morphologies and lipid production were used throughout this study. ICS1, ICS46 and ICS48 were isolated in Bath, UK from blackberry fruits. NCYC2580 was sourced from the National Collection of Yeast Cultures. 4x3 and F3 were strains derived via adaptive evolution of an NCYC2580 progenitor. Strains were maintained on malt extract agar (MEA) plates, and re-streaked on a fortnightly basis. For the preparation of overnight cultures, a single colony was inoculated into 5 mL SMB pH 5 (3% tryptic soy broth, 2.5% malt extract), and incubated at 25 °C with 200 rpm agitation. Optical densities were measured at 595_{nm}. Nitrogen Limited Medium (NLB) was prepared as follows - Glucose 40 g/L, $(NH_4)_2SO_2$ 2 g/L, KH_2PO_4 7 g/L, $MgSO_4$ $7H_2O$ 1.5 g/L, $NaHPO_2$ 2 g/L and yeast extract 1 g/L. This media was autoclaved without glucose or $MgSO_4$, which was added separately after autoclaving individual stock solutions. YNB medium was prepared as follows - 1.78 g/L Yeast Nitrogen Base w/o amino acids, 25 g/L glucose and variable amounts of $(NH_4)_2SO_2$ per desired concentration.

2.3.3. Screening experiments

To screen for lipid production, overnight cultures were prepared as previously described, washed with PBS and diluted to an OD ~1. 500 mL of this diluted culture was used to inoculate 10 mL NLB or YNB depending on experiment. Cultures were incubated for seven days at 25 °C with 200 rpm agitation.

2.3.4. Lipid extraction

The lipid extraction methodology is based upon that proposed by Bligh and Dyer [14], and modified to the following:

Upon completion of culturing, 9 mL of cells were centrifuged at 11.3 thousand RCF, resuspended in 1 mL PBS, and centrifuged again, discarding the supernatant and freezing cell pellets instantly in liquid nitrogen. Following this, frozen cell pellets were freeze dried at -40°C for a minimum of six hours. For the cell disruption, a preweighed amount of cell pellet (ideally within the range of 10 – 100 mg) was mixed with 10 mL 6M HCl, and stirred at 80 °C for one hour. To extract lipids, 10 mL of chloroform:methanol (1:1) was added, and the mixture was stirred overnight at room temperature. To quantify the lipid weight, the lower chloroform phase was removed by hand with a glass pipette, avoiding the emulsion layer which forms between the chloroform and aqueous phase. The chloroform was then fully evaporated via rotary evaporator at 50 °C, and the remaining lipids were weighed to determine the lipid weight % of the cells.

2.3.5. Nile red assay

Nile red staining and assay protocol was an optimised version of that published by Sitepu [2], briefly:

Cells were harvested after seven days and an aliquot of cells was adjusted with PBS to an OD ~1. This dilution was then spun and cells esuspended in fresh experimental media. For the standard curve, glycerol standard solution (2.5 mg/ml equivalent triolein concentration) and isopropanol, with a combined volume of 20 Lwas added to 200 µL of blank experimental media in triplicate to a black 96-well plate. The

standard curve prepared represented a linear range between 20-200 mg/L. For the samples, 200 µL of each diluted culture was added in triplicate to the 96-well plate to which 40 µL Dimethyl sulfoxide (DMSO) was also added. Nile red at a final concentration of 15 µg/mL was added to the standards and samples followed by incubation at 30 °c for 10 minutes. Fluorescence was read with excitation at 530/25, emission 590/35, kinetic readings taken for 20 minutes with 60 second intervals with shaking throughout. Fluorescence values were corrected by cell count as determined by Fast-Read 102 counting slides.

2.3.6. Cell Size Analysis

For image collection, cells were harvested, diluted to OD ~1 using PBS and injected into Fast-read 102 counting slides and imaged using 20x magnification on an EVOS cell imager. For image analysis, GIMP image editor was first used to sharpen and increase the contrast of each image, and the 'Bucket Fill Tool' was used to subtract the background to white [15]. Images were then imported into ImageJ where the image scale was set according to the width of the scale bar in pixels. The image threshold was adjusted, and the 'Analyse particles' tool was used to measure individual cell areas with the following specifications:

- Maximum and minimum size was set from '10-Infinity' to exclude any non-cell debris from the analysis.
- Circularity was set from '0.75-1.00' to exclude instances where cells are in contact with each other.
- 'Exclude on edges' was selected to omit cells cropped by the edges of the image.
- 'Include holes' was selected to only measure the perimeter of the cells, rather than internal areas created by lipid droplets.

2.4. Results.

2.4.1. Nile red optimisation.

Starting from the Nile red staining protocol published by Sitepu [2], a series of optimisation steps were performed with different strains of *M. pulcherrima*, which incorporated high and low lipid acumulators based on mass ratio measurements from traditional extraction. To improve normalisation of fluorescence measurements, the ratio between optical density and cell count for four strains of *M. pulcherrima* was compared (Fig. 1). 4x3 was found to have a cell count substantially lower than the highest strain ICS1 (1.03×10^7 and 9.05×10^7 respectively) when diluted to an OD 1 ± 0.05. The cell count of DH5 and NCYC2580 was higher than that of 4x3, though still far below the count for ICS1.

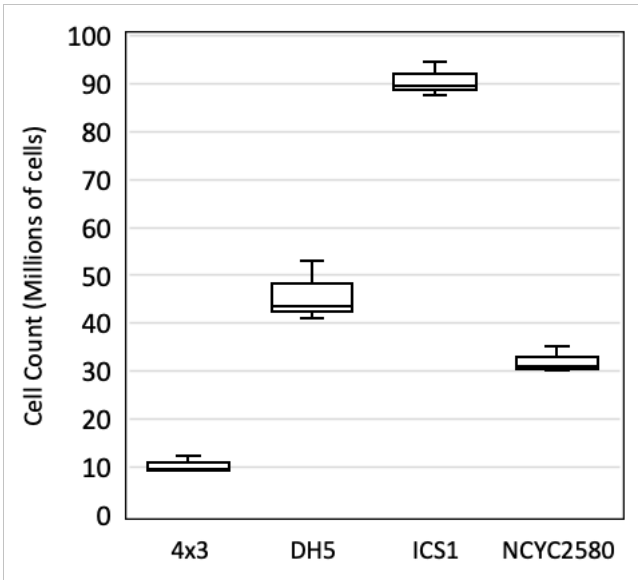


Fig. 1. Cell count of four *M. pulcherrima* strains at the same optical density. Cultures were incubated for seven days in NLB at 25 °C, before dilution to an OD of 1 ± 0.05. Box plots represent triplicate cultures.

Optimisation of DMSO and cell concentration was performed using 4x3, the highest lipid producing strain (Fig. 2). A clear correlation was present between increasing fluorescence measurement and increasing volumes of DMSO added, with each sample achieving maximum fluorescence at 40 µL DMSO. Cellular concentrations between OD 1 and 2 gave the highest values, indicating this as an optimum range. Increasing or decreasing from this range decreased the fluorescence response.

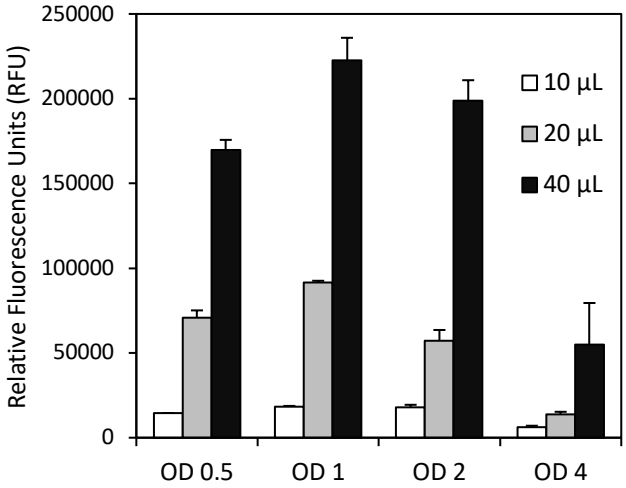


Fig. 2. Effect of cell and DMSO concentration on fluorescence. Triplicate cultures were incubated for seven days in NLB at 25 °C before samples were taken and diluted to the described OD (± 0.05). Neat DMSO was added at the volumes indicated before addition of Nile red. Blank media was added to correct for the dilution effect of higher DMSO volumes. Error bars represent the standard deviation.

Finally, using the optimised cell and DMSO concentrations, different concentrations of Nile red were assayed (Fig. 3). Here, a final concentration of 15 µg/mL achieved the highest reading; increasing or decreasing from this resulted in a lower fluorescence measurement.

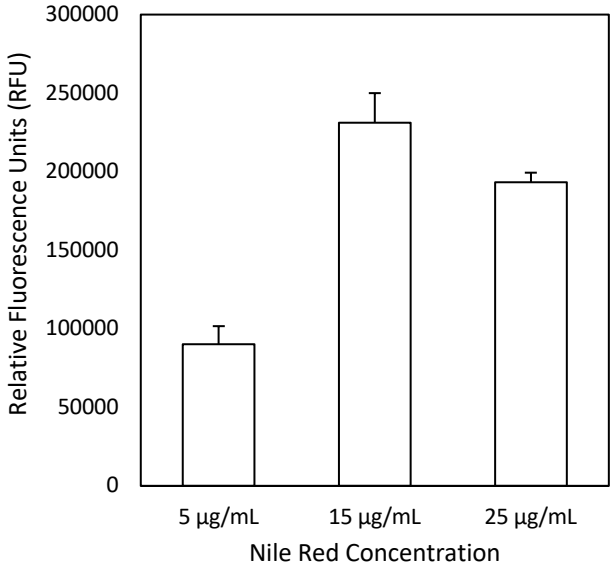
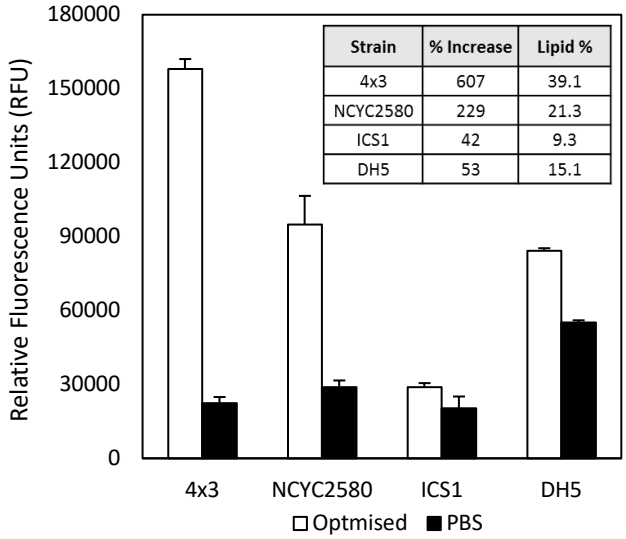


Fig. 3. Effect of Nile red concentration on fluorescence. Triplicate 4x3 cultures were incubated in NLB for seven days in NLB at 25 °C before samples were taken and diluted to an OD 1 (± 0.05). 40 µL DMSO was added as previously determined. Error bars represent the standard deviation.

The optimised protocol was then used to assay four *M. pulcherrima* strains selected to incorporate different cell morphologies and lipid accumulating ranges (Fig. 4). Of the strains used for the optimisation, 4x3, had the greatest fluorescence percentage increase using the optimised method, whereas the increase to ICS1 was the lowest of the strains assayed. Interestingly, the fluorescence percentage increase correlates with the lipid percentage accumulated for each strain, with high accumulating strains 4x3 and NCYC2580 benefiting the most, and low producing strains ICS1 and DH5 benefiting the least.



Strain	% Increase	Lipid %
4x3	607	39.1
NCYC2580	229	21.3
ICS1	42	9.3
DH5	53	15.1

Fig. 4. Comparison of the optimised Nile red method between *M. pulcherrima* strains. Triplicate cultures were incubated in NLB for seven days at 25 °C, before samples were taken and diluted to an OD 1 (± 0.05). Error bars represent the standard deviation. Data included within the box shows the percentage fluorescence increase as a result of the optimised method compared to the equivalent volumes of PBS added during the Nile Red assay, as well as the lipid weight percent for each strain as determined by lipid extraction.

2.4.2. Comparison of Nile red assay and Cell Size Analysis to lipid extractions.

2.4.2.1. Screening multiple strains.

Screening novel strains for lipid production is a common experiment requiring a reliable high throughput screening method. Here, six strains were assayed using both the optimised Nile red method and the cell size analysis (CSA) method developed

here, with results plotted against lipid extraction data from traditional extractions (Fig. 5).

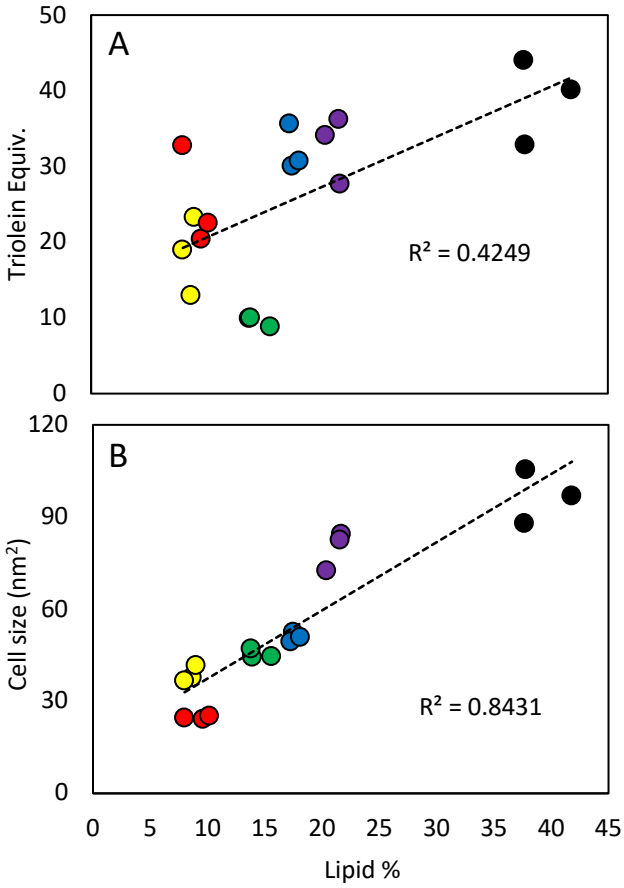


Fig. 5. Comparison of the Nile red assay and CSA for estimating lipid production in six *M. pulcherrima* strains. Triplicate cultures were incubated for seven days in NLB at 25 °C before samples were taken for assays. Graphs represent Triolein equivalents (A) and cell size (B) plotted against lipid extraction values for each sample. Strains and colours: 4x3 – Black, NCYC2580 – Purple, F3 – Blue, ICS46 – Green, ICS1 – Red, ICS48 – Yellow.

A correlation is present in both data sets, however a higher R^2 from the CSA method shows a better representation of lipid extraction data. ICS1 and ICS46, which accumulate <15% lipid, correlate poorly with lipid extractions when assayed with Nile red compared to CSA, which represents their lipid production values accurately. In general, variation within triplicate samples are greater using the Nile red method than CSA. Micrographs taken of these cultures highlight the strain to strain variation in cell morphology, even when grown under the same conditions (Fig. 6).

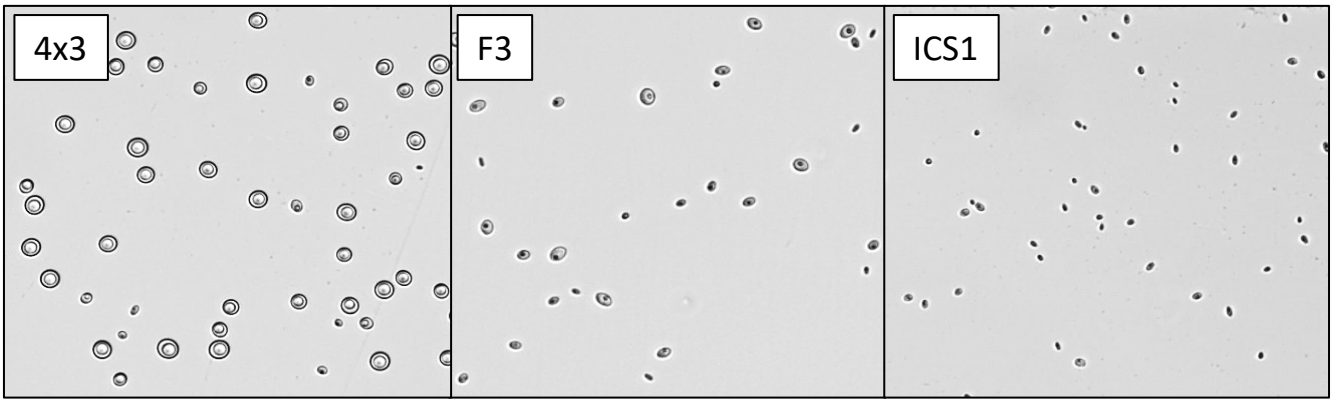


Fig. 6. Micrographs of 4x3, F3 and ICS1. Triplicate cultures were incubated for seven days in NLB at 25 °C before samples were taken for assays. Figure shows a representative section of an image taken with an EVOS cell imager under 20x magnification, with scale bar omitted for clarity.

2.4.2.2. Screening of three strains in two medias

High throughput screening methods are also used to assess the effect changing media types or growth conditions has on lipid accumulation. Using three strains (4x3, NCYC2580 and F3) under high and low nitrogen conditions, both assay methods were again screened for their correlation with traditional lipid extractions (Fig. 7). A strong correlation (R^2=0.92) between CSA and lipid accumulation was achieved, whereas a substantially weaker correlation was seen between Nile red and lipid accumulation (R^2=0.46). Underestimation of lipid production in strains NCYC2580 and F3 grown in low nitrogen conditions by the Nile red assay incorrectly place triolein equivalent values similar to the high nitrogen conditions, despite lipid production decreasing by over half under these conditions. The Nile red assay for strain 4x3 also suggests little difference between both conditions, despite significant differences in actual lipid production. Micrographs of the cultures used for CSA highlight the effect varying nitrogen concentration has on morphology and lipid production, in particular 4x3 which has a reduced lipid droplet size and cell wall thickness when grown in high nitrogen conditions (Fig. 8).

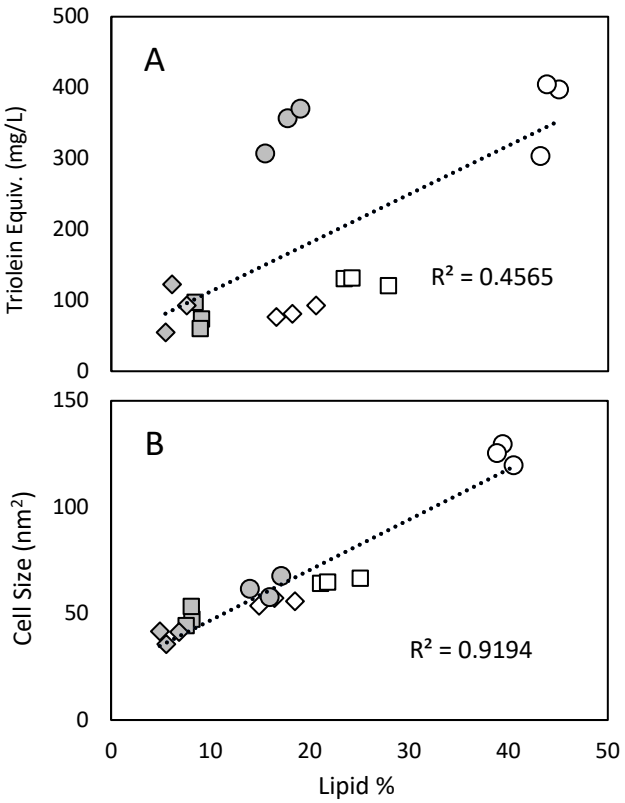


Fig. 7. Comparison of 4x3, F3 and NCYC2580 in high and low nitrogen media. Triplicate cultures were incubated for seven days in YNB at 25 °C before samples were taken for assays. Sole nitrogen source was ammonium sulphate, with high (5.5 g/L) and low (0.3 g/L) nitrogen cultures represented by grey and white markers respectively. Graphs represent triolein equivalents (A) and cell size (B) plotted against lipid extraction values for each sample. Strains: 4x3 – circles, NCYC2580 – squares, F3 – diamonds.

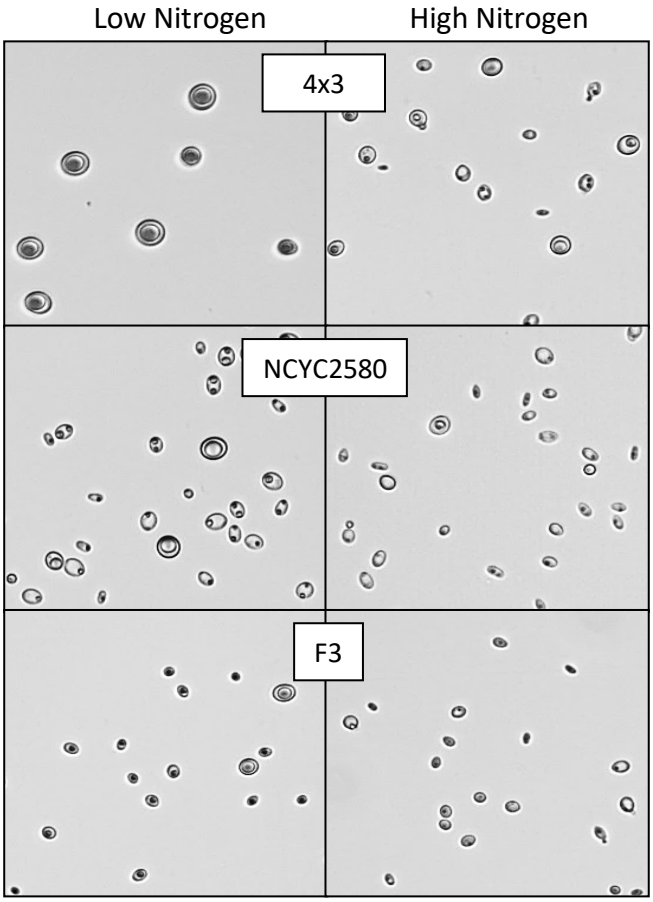


Fig. 8. Micrographs of 4x3, NCYC2580 and F3. Triplicate cultures were incubated for seven days in YNB at 25 °C before samples were taken for assays. High (5.5 g/L) and low (0.3 g/L) nitrogen cultures were created using ammonium sulphate as the sole nitrogen source. Figure shows a representative section of an image taken with an EVOS cell imager under 20x magnification, with scale bar omitted for clarity.

2.4.2.3. Lipid assessment in a single culture over time

Finally, both assay methods were used to track the lipid production of a single culture over time. Samples were taken initially at days 1, 2, 5, 10 and 15. At day 15, a 10x stock of NLB was added to the culture to bring nutrient concentrations back to inoculum levels. Three further samples were taken post addition, representing day 16, 17 and 20. Though only a single culture was performed, robustness was assessed for each method by performing triplicate lipid extractions, gathering cell size data from three picture sets and running triplicate Nile red samples with a fresh triolein standard curve for every time point. CSA accurately tracked the initial increase in lipid production from day 1 to 5, as well as the minor fluctuations in subsequent samples

(Fig. 9). The Nile red method too correlated reasonably well with the initial increase in lipid, however fluctuated significantly across later samples. Each method produced little variation at each timepoint suggesting technically robust methods.

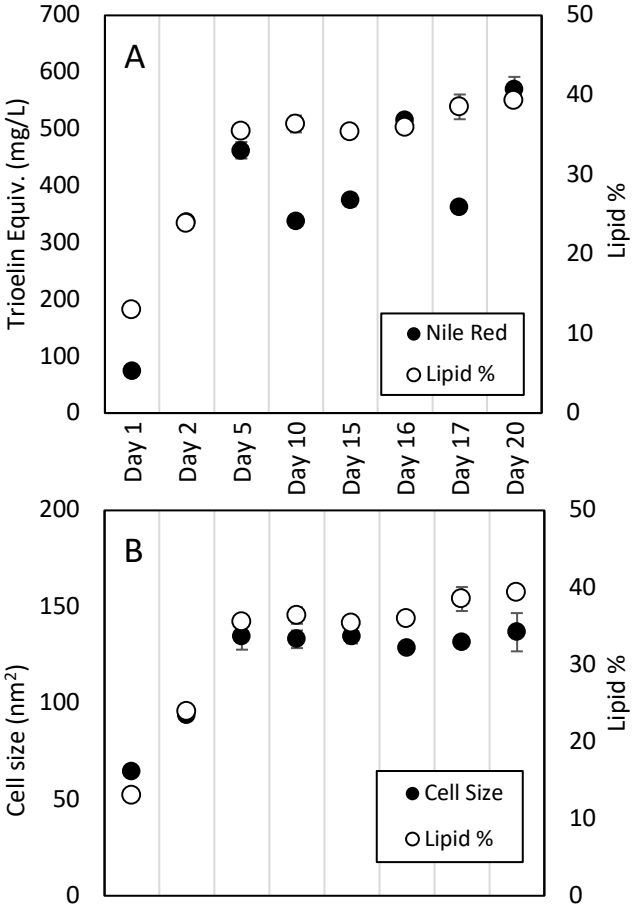


Fig. 9. Comparison of Nile red and cell size analysis when tracking a single culture of 4x3. A single 300 mL culture was performed in NLB for 20 days at 25°C. 10 mL of culture was removed for all three sample methods at the indicated days. At day 15, a 10 mL of a 10x NLB nutrient shot containing the same concentration of starting nutrients was added to the culture. Graphs represent triolein equivalents (A) and cell size (B) plotted against lipid extraction values for each sample. Error bars represent three technical replicates for each timepoint, horizontal scale bar is applied to both parts.

2.4.3. Cell size population dynamics

Previously, the effect of nitrogen concentration on *M. pulcherrima's* ability to accumulate lipid was screened (Fig. 7 and 8). Plotting the data used to calculate an average cell size as a histogram enables additional information about the population dynamics and cell morphology to be determined (Fig. 10). For example, when grown

in low nitrogen conditions, two cell morphologies exist; a group of larger cells clustering around 150 nm^2 and a smaller subset of cells clustering around 50 nm^2. Conversely, only the 50 nm^2 cell type exists within the culture when in the presence of high nitrogen conditions. As these two cell types appear inherently linked to nitrogen concentration as a variable, this additional data provides additional insight into population dynamics.

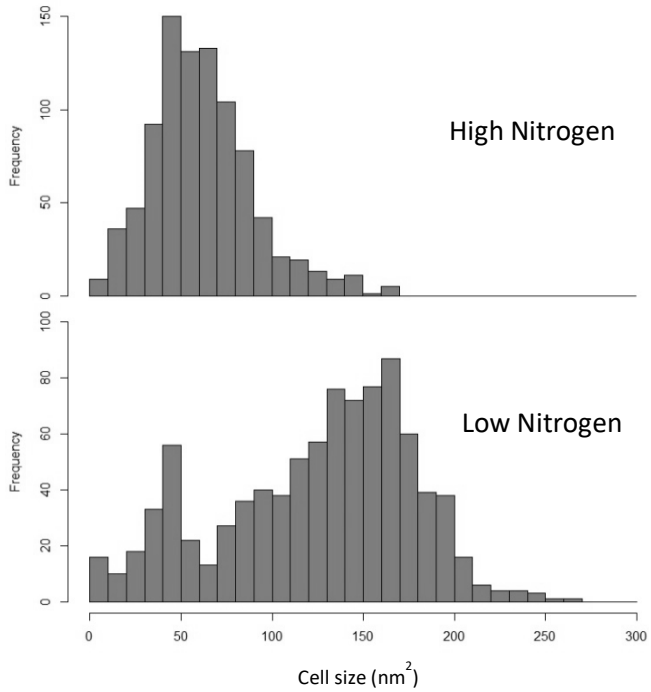


Fig. 10. Histograms plotted of 4x3 cell size data in high and low nitrogen conditions. Triplicate cultures were incubated for seven days in YNB at 25 °C before samples were taken for assays. High (5.5 g/L) and low (0.3 g/L) nitrogen cultures were created using ammonium sulphate as the sole nitrogen source. Histogram plots the data from 900 cells, represented equally from triplicate cultures. Image taken with an EVOS cell imager under 20x magnification.

In addition to understanding the effect media type has on population dynamics, plotting the cell size data of a culture over time can also reveal additional insight about the biological process as it progresses, and how it responds to changes. When displayed as a histogram, the data from Figure 9 shows the population increasing in average cell size between day one, two and five, with the population showing a normal distribution around the average value (Fig. 11). This population dynamic continues until day 16, where the addition of nutrients at day 15 notably splits the population, despite the overall lipid % of the culture showing no change. From day

16 to the final sample at day 20, the lipid production increases from 36% to 39.3%; correlating with the increased number of cells >200 nm^2. Interestingly, during this period the cell population remains split rather than reverting to a normal distribution as observed in the pre-nutrient shot samples.

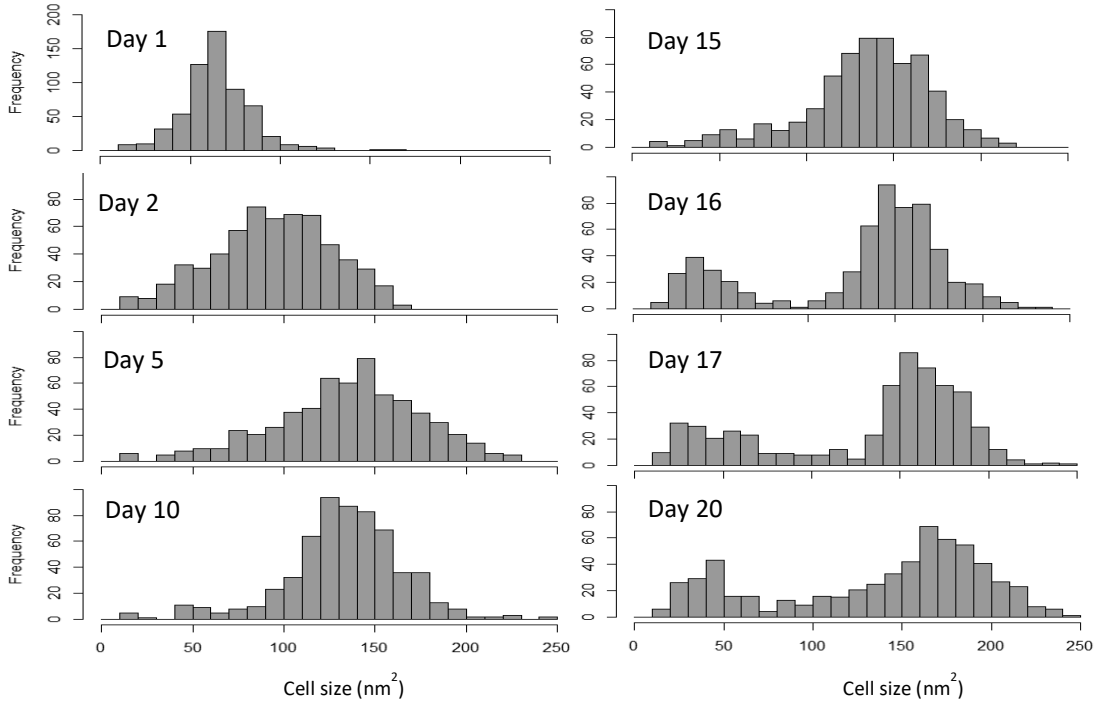


Fig. 11. Histograms plotted of 4x3 cell size data across 20 days of growth. A single 300 mL culture was performed in NLB for 20 days at 25°C. Histogram plots the data from 900 cells, representing pictures from three separately prepared dilutions. Image taken with an EVOS cell imager under 20x magnification.

2.5. Discussion

It is reported that species-specific optimisation of the Nile red assay methodology is a pre-requisite for reliable data acquisition within yeast and microalgae [2,4]. Here, starting with the method published by Sitepu, a Nile red assay was optimised for *M. pulcherrima* [2]. To begin, fluorescence data normalisation was investigated by comparing the OD to cell count ratio between strains. In what is an often overlooked factor, OD was found to be an inaccurate way of normalising data due to the vastly different cell counts present when assayed at the same OD [16]. The most notable demonstration of this is seen between strains 4x3 and ICS1, which when diluted to an OD of 1 give cell counts of 1.03 $x10^7$ and 9.05$x10^7$ respectively. In this instance,

not only does the larger relative size of 4x3 to ICS1 increase light scattering, but the large lipid droplet present is also likely to contribute to this discrepancy. It is important to note that the cell count to OD ratio is not just inconsistent between strains, but also within a strain, as seen by comparing the micrographs and histograms of 4x3 when grown in high and low nitrogen media. Although the requirement for cell counting adds a significant amount of time to the assay, it appears in this case essential for accurate data normalisation.

During the optimisation process, several assay parameters were investigated using the highest lipid accumulating strain, 4x3, with the rationale being that the large lipid droplet and thick cell wall would make this strain the least permeable to Nile red staining. The addition of DMSO to increase permeability is common within most reported protocols, often diluted 1:1 with PBS or culture medium prior to addition [2,17]. Here, the addition of 40 µL undiluted DMSO to 200 µL cells was required to reach the highest fluorescence; far more than the method published by Sitepu which added 25 µL of DMSO/Culture medium (1:1, v/v) to 250 µL sample volume. The addition of 10 µL DMSO within this study, a comparable amount to methods published by Sitepu and Roston, gave a fluorescence reading markedly lower than 40 µL, suggesting *M. pulcherrima* is less permeable to Nile red than those reported elsewhere. A final stain concentration of 15 µg/mL gave the highest measurement, which is again higher than the concentrations published elsewhere. Increasing from this optimum concentration did not however yield a higher measurement, as was observed by Sitepu [2].

Applying the optimised Nile red assay to different *M. pulcherrima* strains uncovered large variation in staining efficacy within this species. Interestingly, although all strains tested had increased fluorescence using the optimised method, this varied considerably between strains. Here, high lipid accumulating strains such as 4x3 and NCYC2580 benefited more from the optimised method than low accumulating strains. Reasons for this again appears attributed to elements of cell morphology, in particular the frequency of the 'pulcherrima' cell type; characterised as a large cell with a thick cell wall, containing a lipid droplet occupying almost the entire cell space

[18]. Due to the frequency of these cell types being highly susceptible to culturing conditions (such as nitrogen availability), it again reiterates the difficulties of optimising staining protocols even to strains within a species.

The second half of this study compared the effectiveness of the optimised Nile red assay with CSA to estimate lipid accumulation. A high throughput screen does not necessarily have to be sophisticated enough to provide a lipid weight % result. Rather, the ability to accurately compare samples within a single experiment and compare these results to previous or future experiments, is an essential. With robustness key to comparing experiments, despite best efforts, it is inevitable that a fluorescence based assay contains more experimental variation than cell imaging. Although technical elements such as dye/standard curve preparation and variation in assay preparation length due to sample number can be kept as consistent as practically possible, dye-specific elements cannot be controlled. Fluorescence quenching, arising from sub-optimal ratios between DMSO, cellular lipid content and dye concentration, is reported to be species specific and likely to be a feature here, particularly in low producing strains given the method was optimised for the highest producing strain [19,20].

In all three comparison studies, CSA was shown to estimate lipid production better than the 96-well plate Nile red assay. CSA performed particularly well in tracking the lipid production of a single culture; an essential assay requirement, demanding robustness and consistency. Whilst the optimised Nile red method requires sample density measurement and dilution, stain preparation, standard curve preparation, preparation and incubation of the microplate followed by a 20-minute time course measurement, CSA needs only sample density measurement and dilution followed by image collection. Reducing the number of steps within a method can therefore reduce errors; improving the data set. Furthermore, the population dynamics gathered by plotting cell size data provides a level of analysis beyond that possible by spectrofluorometric methods alone, using data already collected. This therefore allows real time assessment of the cellular response to changes culturing conditions or nutrient availability.

2.6. Conclusion

The use of Nile red as a high-throughput method to estimate lipid production is common within the study of oleaginous microorganisms. Although this approach can be successful when an optimised method is established, its effectiveness is limited when applying the same methodology to a new species or strain, leading to incorrect results. Here, a Nile red method optimised with the highest lipid producing *M. pulcherrima* strain and a fluorescence-free, image based analysis method developed within this study were compared to traditional lipid extractions for accuracy across three experimental set ups: screening multiple strains, screening a subset of strains in two conditions and tracking a single culture over time. In all instances, CSA had greater correlation with extraction values than Nile red. A key factor inhibiting the success of the optimised Nile red method was the broad range of cell morphologies present within the strains and cultures tested, with cell wall thickness and lipid droplet size likely to have the greatest influence. It is therefore extremely difficult to accurately use a fluorescence based method when variable cell morphology within a strain/species is such a frequent and impacting factor. In these instances, fluorescence free methods such as that presented here are far more effective, and could be of great use in the study of other oleaginous organisms.

Acknowledgements

Thanks go to Dr Deborah Gore-Lloyd, Mauro Moreno, Felix Abeln and Dr. Jim Caunt for their helpful discussions regarding the development of this project. This research has been funded by the Industrial Biotechnology Catalyst (Innovate UK, BBSRC, EPSRC) to support the translation, development and commercialisation of innovative Industrial Biotechnology processes (EP/N013522/1) and by the EPSRC through the Centre for Doctoral Training in Sustainable Chemical Technologies (EP/L016354/1).

Competing interests

The author(s) declare that they have no competing interests.

2.7. References

[1] Whiffin, F.; Santomauro, F.; Chuck, C.J. Toward a microbial palm oil substitute: Oleaginous yeasts cultured on lignocellulose. *Biofuels, Bioprod. Biorefining* 2016.

[2] Sitepu, I.R.; Ignatia, L.; Franz, A.K.; Wong, D.M.; Faulina, S.A.; Tsui, M.; Kanti, A.; Boundy-Mills, K. An improved high-throughput Nile red fluorescence assay for estimating intracellular lipids in a variety of yeast species. *J. Microbiol. Methods* 2012, *91*, 321–328.

[3] Orr, V.; Rehmann, L. Improvement of the Nile Red fluorescence assay for determination of total lipid content in microalgae independent of chlorophyll content. *J. Appl. Phycol.* 2014, *27*, 2181–2189.

[4] Rumin, J.; Bonnefond, H.; Saint-Jean, B.; Rouxel, C.; Sciandra, A.; Bernard, O.; Cadoret, J.-P.; Bougaran, G. The use of fluorescent Nile red and BODIPY for lipid measurement in microalgae. *Biotechnol. Biofuels* 2015, *8*, 42.

[5] Smith, A.E.; Zhang, Z.; Thomas, C.R.; Moxham, K.E.; Middelberg, AP. The mechanical properties of Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. U. S. A.* 2000, *97*, 9871–4.

[6] Sandager, L.; Gustavsson, M.H.; Ståhl, U.; Dahlqvist, A.; Wiberg, E.; Banas, A.; Lenman, M.; Ronne, H.; Stymne, S. Storage lipid synthesis is non-essential in yeast. *J. Biol. Chem.* 2002, *277*, 6478–6482.

[7] Doan, T.T.Y.; Obbard, J.P. Improved Nile Red staining of Nannochloropsis sp. *J. Appl. Phycol.* 2011, *23*, 895–901.

[8] Balduyck, L.; Veryser, C.; Goiris, K.; Bruneel, C.; Muylaert, K.; Foubert, I. Optimization of a Nile Red method for rapid lipid determination in autotrophic, marine microalgae is species dependent. *J. Microbiol. Methods* 2015, *118*, 152–158.

[9] Raschke, D.; Knorr, D. Rapid monitoring of cell size, vitality and lipid droplet development in the oleaginous yeast Waltomyces lipofer. *J. Microbiol. Methods* 2009, *79*, 178–183.

[10] Capus, A.; Monnerat, M.; Ribeiro, L.C.; de Souza, W.; Martins, J.L.; Sant'Anna, C. Application of high-content image analysis for quantitatively estimating lipid accumulation in oleaginous yeasts with potential for use in biodiesel production. *Bioresour. Technol.* 2016, *203*, 309–317.

[11] Gao, C.; Xiong, W.; Zhang, Y.; Yuan, W.; Wu, Q. Rapid quantitation of lipid in microalgae by time-domain nuclear magnetic resonance. *J. Microbiol. Methods* 2008, *75*, 437–440.

[12] Ami, D.; Posteri, R.; Mereghetti, P.; Porro, D.; Doglia, S.M.; Branduardi, P. Fourier transform infrared spectroscopy as a method to study lipid accumulation in oleaginous yeasts. *Biotechnol. Biofuels* 2014, *7*, 12.

[13] Huang, Y.Y.; Beal, C.M.; Cai, W.W.; Ruoff, R.S.; Terentjev, E.M. Micro-Raman spectroscopy of algae: Composition analysis and fluorescence background behavior. *Biotechnol. Bioeng.* 2010, *105*, 889–898.

[14] Bligh, E.G.; Dyer, W.J. A rapid method of total lipid extraction and purification. *Can. J. Biochem. Physiol.* 1959, *37*, 911–917.

[15] Kimball, S.; Mattis, P. GIMP. *GNU Image Manip. Progr.* 1996.

[16] Sutton, S. Measurement of microbial cells by optical density. *J. Valid. Techn*

2011, 46–49.
[17] Rostron, K.A.; Lawrence, C.L. Nile red staining of neutral lipids in yeast. In *Methods in Molecular Biology*; 2017; Vol. 1560, pp. 219–229.
[18] Santamauro, F.; Whiffin, F.M.; Scott, R.J.; Chuck, C.J. Low-cost lipid production by an oleaginous yeast cultured in non-sterile conditions using model waste resources. *Biotechnol. Biofuels* 2014, *7*, 34.
[19] Pick, U.; Rachutin-Zalogin, T. Kinetic anomalies in the interactions of Nile red with microalgae. *J. Microbiol. Methods* 2012, *88*, 189–196.
[20] Chen, W.; Sommerfeld, M.; Hu, Q. Microwave-assisted Nile red method for in vivo quantification of neutral lipids in microalgae. *Bioresour. Technol.* 2011, *102*, 135–141.

This declaration concerns the article entitled:										
Comparison of Nile red and cell size analysis for high throughput lipid estimation within oleaginous yeast.										
Publication status (tick one)										
draft manuscript		**Submitted**		**In review**	X	**Accepted**		**Published**		
Publication details (reference)										
Candidate's contribution to the paper (detailed, and also given as a percentage).	The candidate contributed to/ considerably contributed to/predominantly executed the… Formulation of ideas: The phenomena of cell size as an indication of lipid production was discussed between myself and colleagues within the group. Idea to develop this into a high throughput lipid quantification method allowing for the evaluation of population dynamics was my own idea. 100% Design of methodology: Methodology was designed in its entirety by myself. 100% Experimental work: Experimental work was completed in its entirety by myself. 100% Presentation of data in journal format: Formatted for: European Journal of Lipid Science and Technology.									
Statement from Candidate	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.									
Signed							**Date**			

3. Optimisation of semi continuous bioprocessing through cell dynamics monitoring.

3.1. Commentary

In the previous chapter, an accurate and robust lipid estimation method was developed which also allowed per-cell population dynamics of lipid accumulation to be analysed.

During method validation, an interesting phenotype was uncovered which could have possible implications on a scaled up industrial bioprocess. In brief, when cells which had accumulated lipid were transferred into fresh media, or given a nutrient shot, a sub-population of the cells lost the oleaginous phenotype. This observation was seen elsewhere within the group in bioreactor studies and represents a potential issue for continuous or semi-continuous processing.

By using the method established in chapter one, cell size and population dynamics were investigated in a series of different culturing set ups to elucidate the factors causing this growth phenotype, as well as proposing a solution to this issue.

Optimisation of semi continuous bioprocessing through cell dynamics monitoring.

Authors: Robert H. Hicks[a], Felix Abeln[b], Christopher J. Chuck[b], Roderick J Scott[c] , David J. Leak[c] , Daniel A. Henk[c].

[a]Corresponding author. Centre for Doctoral Training in Sustainable Chemical Technologies, Department of Biology and Biochemistry, University of Bath, Bath, BA2 7AY, United Kingdom. Email: R.H.Hicks@bath.ac.uk

[b] Department of Chemical Engineering, University of Bath, Bath, BA2 7AY, United Kingdom

[c] Department of Biology and Biochemistry, University of Bath, Bath, BA2 7AY, United Kingdom

Key Words: Semi continuous, Microbial oil, Bioprocess, Oleaginous

Abstract.

M. pulcherrima is a promising candidate strain for the production of microbial oil. Though there are many reported advantages for the use of this organism within an industrial process, it has not yet been determined whether this organism is suited to the demands of a continuous or semi continuous process.

Here, it was found that lipid production can be negatively affected when new cells are grown in the presence of supernatant taken from an existing culture. Furthermore, it was shown that the lipid production of new cells was more or less affected depending on the age of the culture from which the supernatant came from. *M. pulcherrima* was then trailed within a semi continuous process, where it was found that continuously culturing the same cell line caused lipid production to plateau after serially transferring cells more than twice. To overcome this production plateau, a semi continuous model was developed whereby freshly grown cells were introduced into the bulk culture at a low frequency at the start of each batch. In addition, a system whereby the main culture was partially harvested and replaced with fresh media proved successful in limiting the possibility of cells remaining in the main culture for more than two transfers. Combining these two processes provided a semi continuous model which was then validated at 250 L scale, where no production plateau was encountered.

3.2. Introduction

The industrial production of microbial oils has great potential to be a sustainable alternative to fossil fuels and vegetable oils. In order to develop a microbial oil bioprocess capable of economically competing with the unsustainable alternatives, low cost lignocellulosic feedstocks are commonly used. The technical challenges faced within lignocellulosic processes are however numerous, and include effective feedstock pre-treatment, conversion of different carbon sources into product, tolerance to fermentation inhibitors, and downstream extraction and processing of products[1]. An often-overlooked aspect of any bioprocess however is the suitability of the platform organisms to continuous or semi continuous culturing. As economic feasibility can be greatly affected by this, it is an important aspect to consider when designing the bioprocess.

Industrial fermentations are most often performed in four ways: batch, fed-batch, continuous or semi-continuous. Batch fermentations consist of a process where all of the feedstock is present at the start of culturing whereas fed-batch fermentations alter slightly in that some nutrients are supplied at pre-determined times during the processes progression[2]. Fed-batch cultures are generally higher yielding than batch, as certain nutrients when present at high concentrations from the start, nitrogen within an oleaginous process for example, can be detrimental to product production. Due to the one-off nature of batch and fed-batch fermentations, these processes can be intensively optimised for yield and productivity, often through the use of experimental design software such as design of experiments. In the same way however, batch or fed-batch processes are only harvested once, meaning the time consuming process of cleaning and re-sterilising has to be carried out prior to the next process run. Batch or fed-batch systems are therefore best suited to the production of high value products such as monoclonal antibodies and are particularly useful when strict process and quality control is required.

The main difference with a continuous or semi-continuous process is that cells or product is continuously, or periodically, harvested across an extended period of time.

The major benefit of this system is that the costly 'downtime' necessary for cleaning and re-sterilising is either removed or significantly reduced. Although continuous processes cannot be optimised to the same extent as fed-batch systems, product formation over a longer time frame is greater as the system is not fluctuating between batch cycles[3]. Complications with continuous culturing can occur if cells are susceptible to product or by-product inhibition[4,5], or if microbial contamination occurs[3]. The organism itself is also likely to adaptively evolve to the conditions within the culture as it is maintained in an environment of defined selective pressure. This is particularly likely within lignocellulosic systems where fermentation inhibitors are present[6]. Though this could be beneficial for the process, for instance if contamination is reduced as the adapted strain can survive 'harsher' environments than potential contaminants, adaptive evolution is unpredictable and could result in significantly reduced product formation[6].

The oleaginous yeast *Metschnikowia pulcherrima* has previously been investigated as a potential platform organism for microbial oil production. For an economically viable process, production would be required in high volume at low cost, meaning a continuous or semi-continuous system is best suited. Through the use of a cell size based method of lipid estimation, previous studies have shown an interesting phenotype when a lipid-rich population of *M. pulcherrima* cells were given additional nutrients[7]. Here, the normal distribution of cell sizes within the existing population pre nutrient addition split, resulting in a fraction of non-oleaginous smaller cell types (typically sized <100 nm^2) which fail to develop into lipid rich cells to become present (Figure 1). Phenotypic effects caused by changes in culturing conditions such as this could be potentially damaging to yield and productivities, especially considering a continuous culture. Using cell size analysis, HPLC and cell population dynamics, various culturing conditions and semi continuous strategies were trailed to elucidate factors affecting *M. pulcherrima's* suitability to a continuous bioprocess.

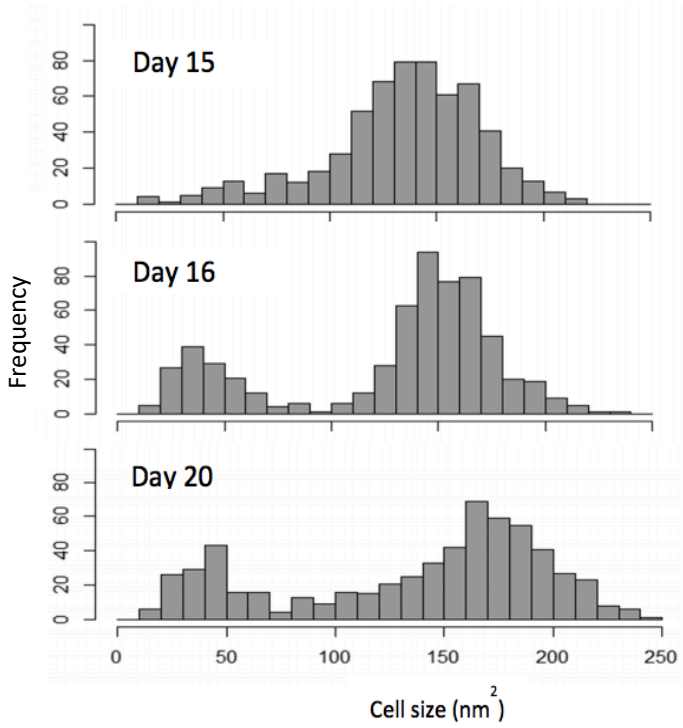


Figure 1. Effect of nutrient addition on cell size population dynamics. Upon nutrient addition at day 15, the cell size population dynamics split from a normally distributed culture, to a spilt distribution where the smaller sized cells did not accumulate lipid.

3.3. Methods.

3.3.1. Chemicals

Unless otherwise stated, chemicals were sourced from Sigma Aldrich and used without further purification.

3.3.2. Strains, strain maintenance and media

The *M. pulcherrima* strain '4x3' used throughout this study was obtained from a previous evolutional experiment within the group, having derived originally from NCYC2580 sourced from the National Collection of Yeast Cultures[6]. Strains were maintained on malt extract agar (MEA) plates, and re-streaked on a fortnightly basis. For the preparation of overnight cultures, a single colony was inoculated into 10 mL SMB pH 5 (3% tryptic soy broth, 2.5% malt extract), and incubated at 25 °C with 200 rpm agitation. Optical densities throughout were measured at 595_{nm}. Media used

within all experiments was a nitrogen limited broth (NLB) consisting of: Carbon source (e.g. xylose, or as described) 40 g/L, $(NH_4)_2SO_2$ 2 g/L, KH_2PO_4 7 g/L, $MgSO_4$ $7H_2O$ 1.5 g/L, $NaHPO_2$ 2 g/L and yeast extract 1 g/L. This media was autoclaved without the carbon source or $MgSO_4$, which was added separately after autoclaving individual stock solutions. Experimental cultures were performed at 25 °C with 200 rpm agitation for the duration specified within figure legends.

For large scale bioreactor culturing, a simplified NLB media was prepared by mixing 5x 25 kg pure liquid glucose syrup (HarryHarvey) and 1.134 kg yeast extract with water to an approximate volume of 250 L. Bioreactor was inoculated with 2 L of an SMB overnight, split into 2x 1 L shake flasks. Culture was harvested and replaced with simplified NLB at regular intervals, as described within the main text. SMB and simplified NLB was prepared throughout as described. Temperature was maintained at approximately 20 °C and stirring maintained at the same level throughout.

3.3.3. Culturing method schematics

The following are schematics to simplify the culturing strategies used throughout the results chapter, with each schematic labelled according to the relevant figure.

Figure 2:

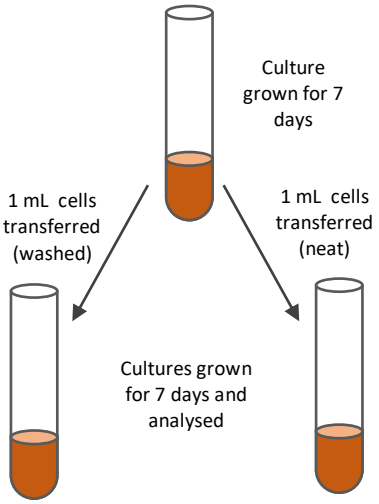


Figure 3:

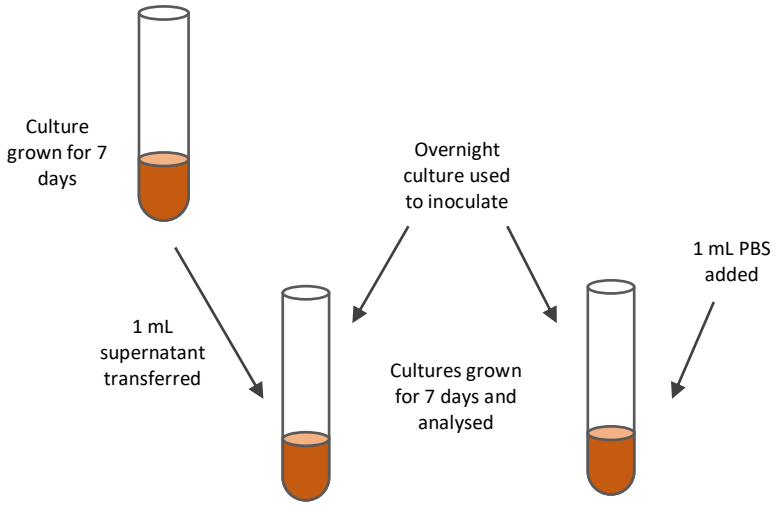


Figure 4:

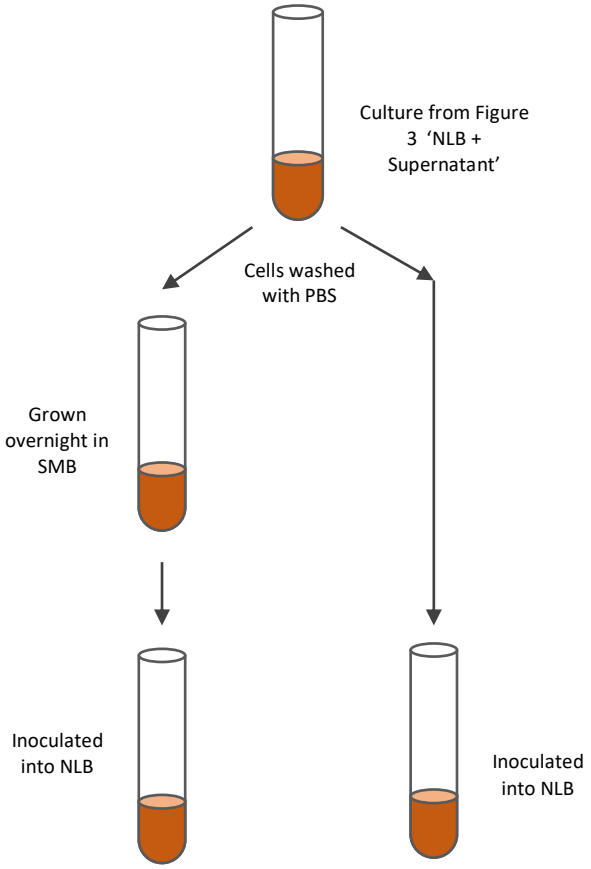


Figure 5:

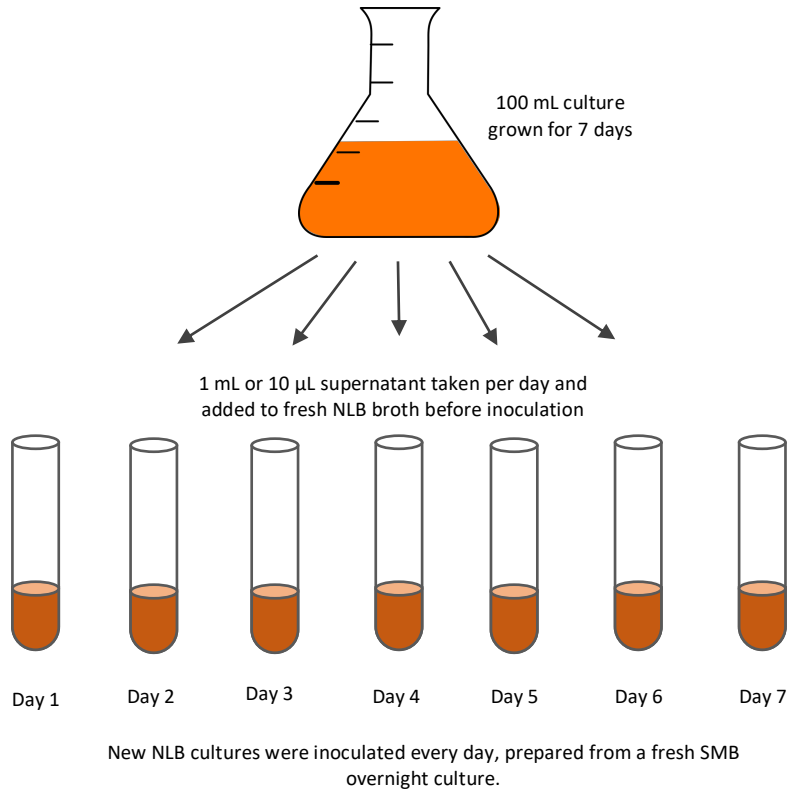


Figure 7:

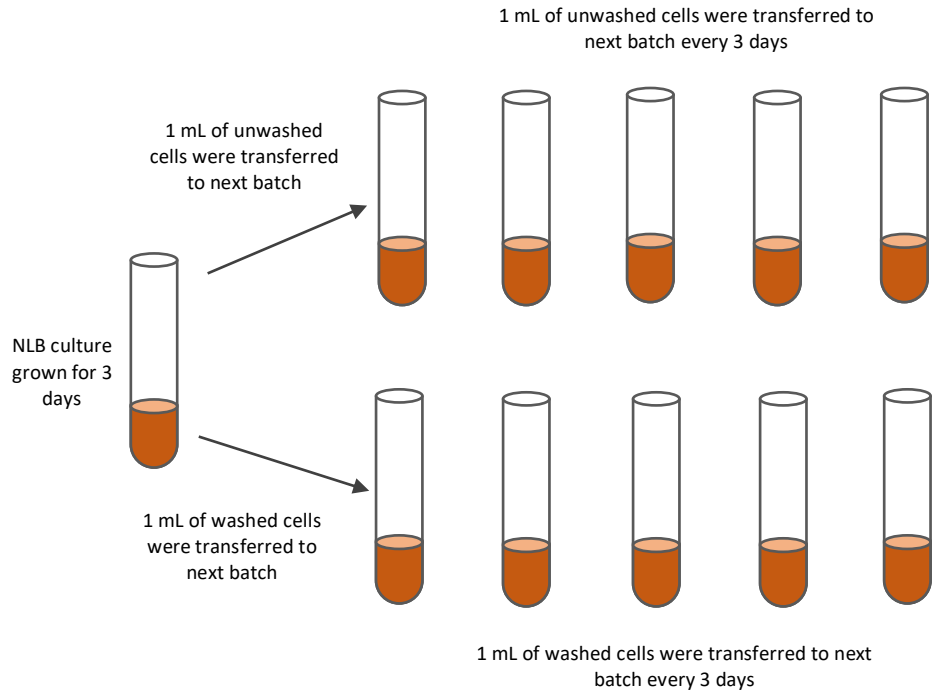
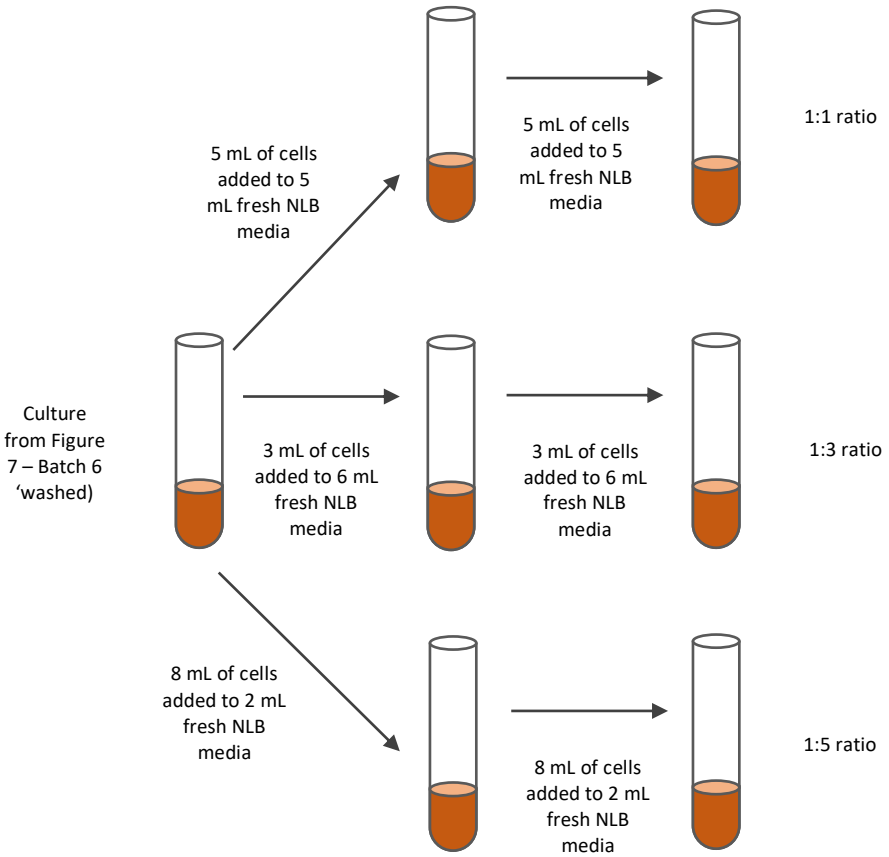


Figure 9:



3.3.4. HPLC

For HPLC analysis, supernatant samples were diluted with ddH_2O where appropriate and filtered (0.22 μm). Glucose, arabinose, ethanol and glycerol content was determined using Agilent 1260 Infinity LC system using an 300 × 7.8 mm Rezex™ RHM-Monosaccharide H+ (Phenomenex) HPLC column held at 80 °C and refractive index detector at 40 °C. A 5-point calibration curve was performed to quantify glucose, while other analytes were quantified by their fluctuating relative peak area.

3.3.5. Cell size analysis

For image collection, cells were harvested, diluted to OD ~1 using PBS and injected into Fast-read 102 counting slides and imaged using 20x magnification on an EVOS cell imager. Full details of image analysis performed as described elsewhere[7].

3.4. Results and Discussion.

3.4.1. Factors affecting population dynamics.

In previous work it was shown that the addition of nutrients to a population of lipid-rich cells caused a split in the cell size of the population; giving rise to smaller, non-oleaginous cell types. To confirm a cell size split within the experimental set up used within this study (transferring cells into fresh media rather than adding a nutrient shot), cells were grown for seven days and then reinoculated into fresh media. To investigate this effect further, cells were either transferred neat, or washed three times with PBS. As before, the population size dynamics splits upon transfer to fresh culture, indicating again the presence of a smaller cell type (< 100 nm^2) unable to accumulate lipid when cultured within a second batch (Figure 2). Though the same population dynamic emerged, a difference was seen between the average cell size and biomass between the cultures inoculated with washed cells versus those transferred directly from the previous culture. From the plotted data, the proportion of small, non-oleaginous cell types appears consistent between the conditions, however the washed cultures have a higher frequency of cells > 200 nm^2, indicating higher lipid accumulation. Both cell groups were inoculated into the same volume of media, meaning the nutrients available were equal. This infers the possibility that something carried over within the supernatant could be having an inhibitory effect on lipid accumulation.

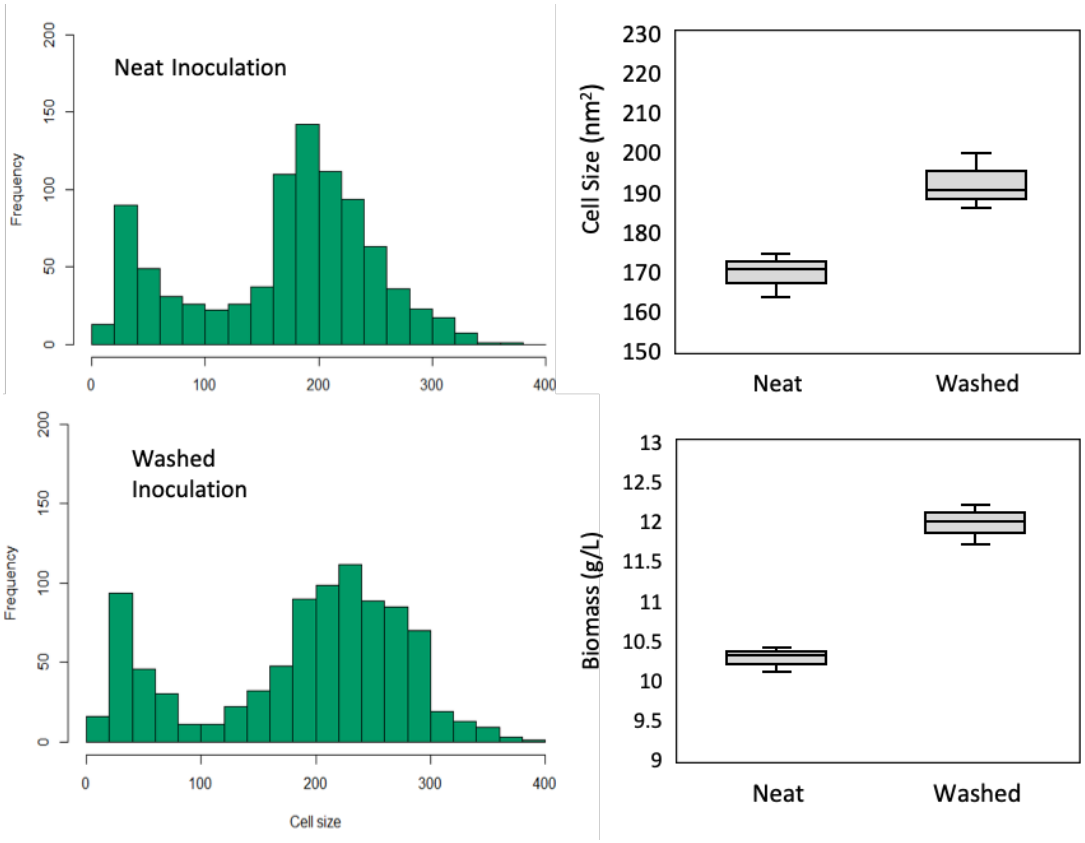


Figure 2. Re-inoculating an established culture into new media. Cells were cultured for seven days in 10 mL of NLB, at which point 6 mL was removed. 3 x 1 mL aliquots were transferred directly into fresh NLB media (termed 'Neat' within figures), and 3 x 1 mL aliquots were washed three times in PBS (termed 'Washed' within figures), before being transferred into fresh media. Cultures were incubated for seven days, upon which cells were imaged and 1 mL was taken for biomass. Box plots represent data from triplicate cultures. Histograms represent data from 900 cells, with 300 cells from each culture used.

To further investigate the potential effect carrying over supernatant from an older culture could have on lipid accumulation, fresh cells from an SMB overnight were inoculated into 10 mL NLB media containing either 1 mL of supernatant from an old culture, or the equivalent volume of PBS. A significant result was obtained, with the cell size and biomass of the NLB + supernatant cultures far below the control cultures (Figure 3). In fact, very few of the cells in the NLB + supernatant cultures generated an oleaginous phenotype; generally classified as > 200 nm^2 within this organism. Though the final cell size was far below the control sample, a final biomass of 2.2 g/L indicates that it wasn't the case that the inoculated cells did not grow.

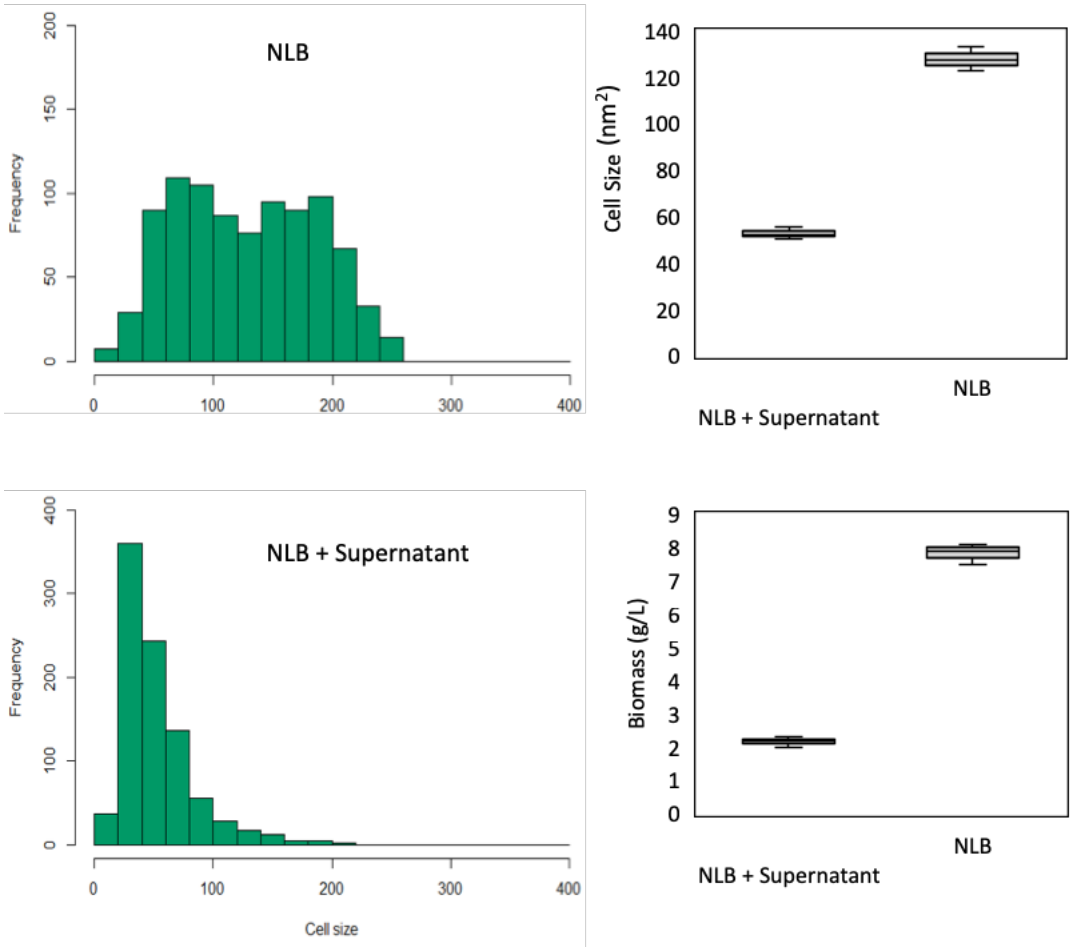


Figure 3. Supernatant effect on new cells. Cells were cultured for seven days in 10 mL of NLB. Media was supplemented with either 1 mL of supernatant from a previously grown seven-day culture, or 1 mL of PBS. Cultures were incubated for seven days, upon which cells were imaged and 1 mL was taken for biomass determination. Box plots represent data from triplicate cultures. Histograms represent data from 900 cells, with 300 cells from each culture used.

Using the cells from Figure 3 which failed to accumulate lipids (NLB + Supernatant), it was tested whether these cells could become oleaginous if washed and subcultured directly into fresh NLB, or if washed, subcultured into SMB, and then inoculated into NLB. Here, in both instances biomass and cell size was greater than the starting culture (Figure 4). In fact, cells which were first subcultured in SMB before being inoculated into NLB returned a cell size and biomass almost akin the control culture in Figure 3. Though this appears a successful strategy, it is unlikely that subculturing the bulk of cells in SMB would be possible within a semi-continuous process. An important result from this experiment however is how malleable the

oleaginous phenotype in this strain of *M. pulcherrima* is with respect to culturing conditions.

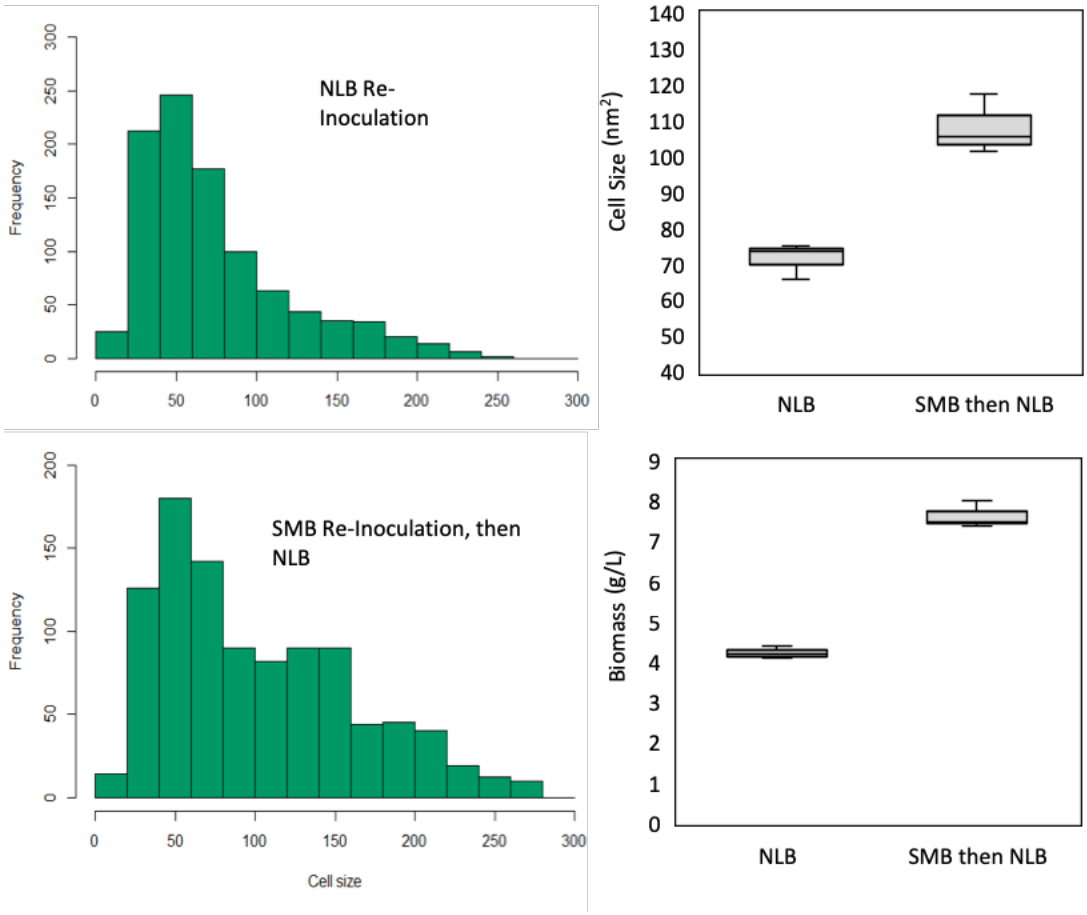


Figure 4. Re-inoculating non-oleaginous cells. 6 mL of cells from the end point of Figure 3 were washed three times with PBS. In triplicate, 1 mL of washed cells were inoculated directly into NLB, or into SMB, grown overnight, and then inoculated into NLB. Histograms represent data from 900 cells, with 300 cells from each culture used.

With evidence that the supernatant from an existing culture was having a deleterious effect on the lipid production of fresh cells, it was next determined whether taking supernatant from different timepoints points of an existing culture had a variable effect. In addition, the effect of supernatant volume was investigated by adding either 1 mL to 10 mL of fresh media as previous, or 10 uL. As this experiment required taking over 3 mL of supernatant per day from the 'stock' culture, a 100 mL shake flask was performed to ensure volume enough was available without altering its volume. This experiment would therefore also test whether the different culturing volume and conditions within the shake flask set versus the culture tube would generate a

different supernatant effect. Contrary to data from Figure 3, the supernatant from the shake flask was not having as detrimental an effect on the lipid production of fresh cells (Figure 5). That said, a varying effect on lipid production was seen as a result of adding supernatant from different timepoints. In particular, adding 1 mL of supernatant from day three of the stock culture yielded a markedly lower cell size than was observed at any other data point. Interestingly, adding supernatant from day two of the stock culture appears to have the reverse effect on lipid accumulation, particularly when adding the lower 10 µL volume. The variability that the supernatant has on the lipid accumulation and population dynamics of new cells is highlighted by comparing the day two and three 10 µL supernatant histograms (Figure 5B).

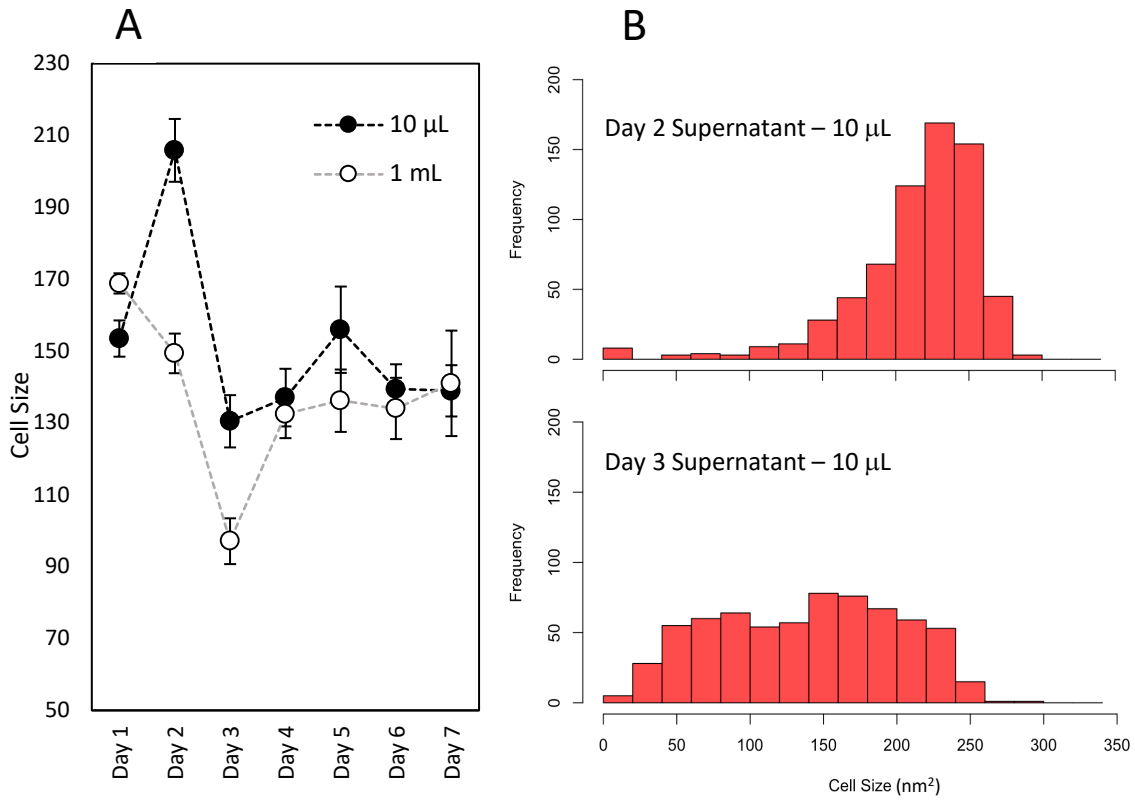


Figure 5. Time-course supernatant effect on the growth of new cultures. A: 100 mL NLB culture was inoculated from an SMB overnight of 4x3. Six-hours post inoculation (Day 1) and then every day from days two to seven, supernatant was removed from this culture and added to fresh 10 mL NLB cultures in either 1 mL or 10 µL volumes. Cells from a freshly prepared SMB overnight were inoculated into these NLB + supernatant culture tubes and cultured for three days. Data points represent the

average cell size within triplicate cultures. B: Data collected for average cell size were plotted as a histogram to display population dynamics.

To further investigate the effect of supernatant on the lipid accumulation of fresh cells, HPLC was performed on the stock culture supernatants prior to addition. Comparing HPLC results with cell size data shows a correlation between decreased cell size in the 'day three' culture with peak ethanol concentration at transfer of supernatant (Figure 6). Increased ethanol concentration within culture medium has previously been linked with a reduction in lipid production, correlating with results found here[8]. Interestingly however, although ethanol concentrations within the supernatant transferred on day two and four were still approximately half that of the day three sample, no negative effects on cell size were seen. It is possible therefore that because only a small population of cells are inoculated into new cultures from the overnight culture, the per-cell dosage of ethanol within the day three supernatant supplemented culture may have been enough to cause this on lipid production. This idea is supported as only the 1 mL supernatant addition caused a notable reduction in cell size.

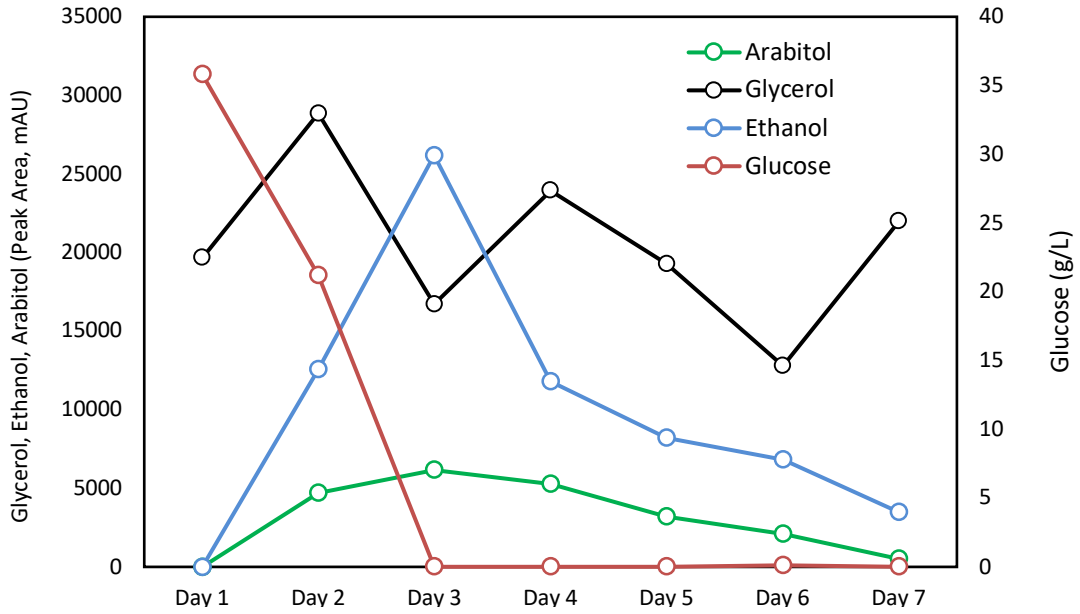


Figure 6. HPLC analysis of time-course supernatant samples. A 100 mL NLB culture was inoculated from an SMB overnight of 4x3. Six-hours post inoculation (Day 1) and then every day, from days two to seven, supernatant was removed from this culture and analysed via HPLC.

From the data available, it isn't possible to hypothesise what caused increased cell size within cultures inoculated into media containing 10 μL supernatant from day two of the stock culture. It is possible with this result, and indeed the result discussed above, that the range of compounds identified via HPLC are not responsible for any of the observed phenotypic effects. For example, an alternative theory could be that cells growing within the stock culture had become glucose-exhausted for the first time by the time the day three supernatant was taken, meaning signalling molecules or proteins related to glucose starvation were present within the media and could potentially influence the growth of fresh cells.

3.4.2. Semi-continuous culturing.

Data thus far has explored the sensitivity of *M. pulcherrima* lipid accumulation with respect to the effects caused by supernatant. Next, the effect of culturing the same cell line continuously for successive batches was evaluated. For this, a semi continuous model was adopted whereby batches were cultured for three days, at which point 1 mL of washed or unwashed cells would be transferred to the next NLB broth, grown for a further three days, before repeating the process further. Though a semi continuous process is unlikely to harvest 90% of the culture as this would represent (by only taking forward 1 mL out of a 10 mL culture), this would still evaluate the ability of *M. pulcherrima* cells to be cultured continuously. By washing the cells or transferring neat, it would also be possible to determine whether any observed effects of continuous culturing could be attributed to carry-over of supernatant components or simply an effect of cell age, for example.

In agreement wither earlier data (Figure 2), the first transfer of cells from an initial culture to a fresh media resulted in a significantly higher cell size (Figure 7A). In addition, washing the cells prior to transfer resulted in increased cell size versus transferring cells neat, again correlating with earlier data. Transferring cells from this second batch to a third however caused a significant reduction in cell size in both washed and neat cultures and average cell sizes were lower than the initial batch in

both cases. Subsequent transfers failed to resurrect the oleaginous phenotype; continuing to reduce until the final batch six. Analysis of the final culture showed an almost complete absence of oleaginous cell types (Figure 7B). Given the same trend is seen regardless of whether cells were washed before transferring, it appears likely that the loss of the oleaginous phenotype is independent from the supernatant effect observed within the earlier study.

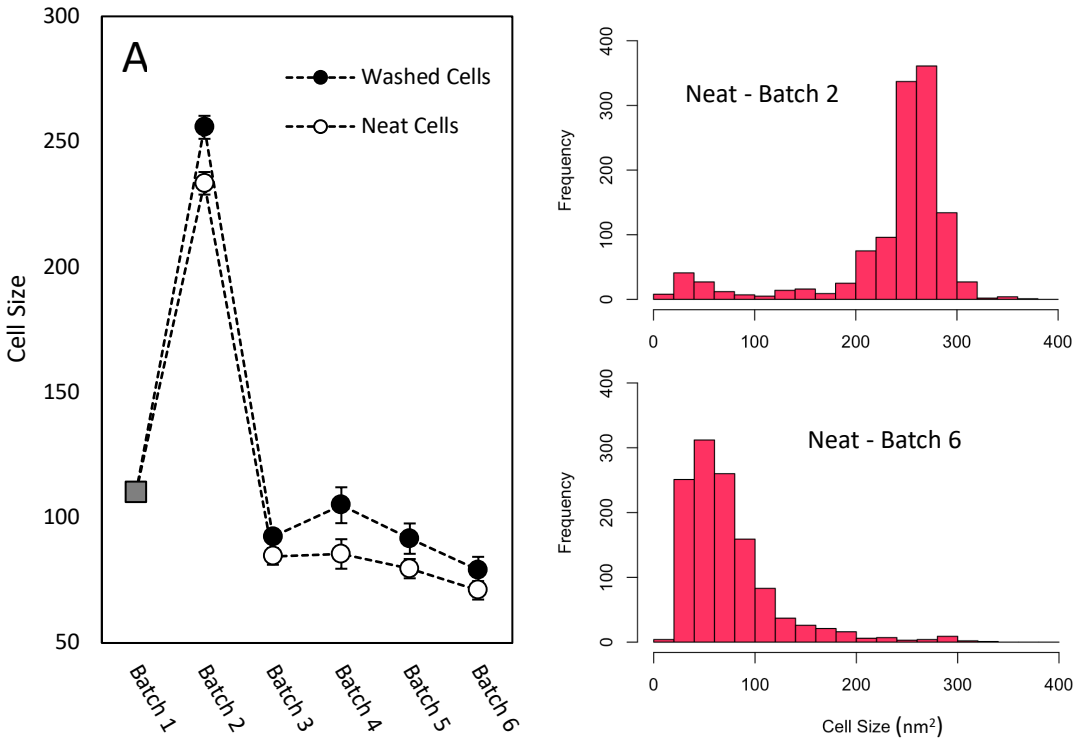


Figure 7. Cell size over time within a semi continuous model. An initial NLB culture was inoculated from a fresh SMB overnight, and grown for three days. Then, 3x 1 mL PBS washed aliquots were used to inoculate 3x fresh NLB cultures and 3x 1 mL unwashed aliquots were transferred neat to 3x fresh NLB broth. This was repeated for successive batches, until the final batch six. A) Average cell size represents triplicate cultures. B) Histograms represent 900 cells; 300 from each culture within the triplicate.

HPLC analysis of the pre-transfer culture supports this theory, as the data shows that none of the analysed compounds are present at high concentrations in the later batches when cell size is at its lowest, relative to earlier batches (Figure 8A & B). One interesting result however was the incomplete glucose consumption within batch two, where the highest lipid accumulation occurred. In both sets of cultures, almost

30 g/L glucose remained from a starting concentration of 40 g/L after three days of culturing. This, together with the high cell size in batch two relative to the initial culture, suggests that the cells transferred are using the available glucose for further lipid accumulation, rather than for replication.

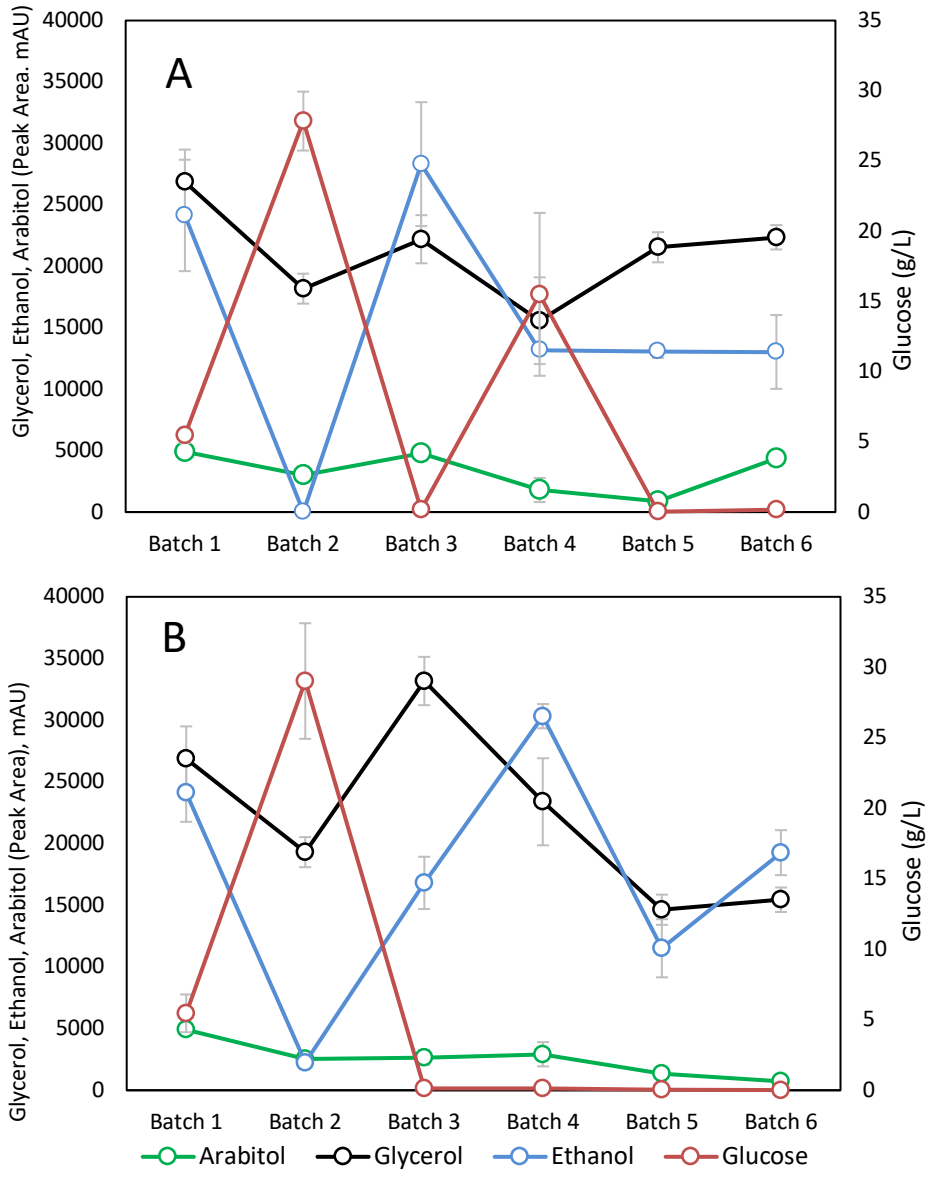


Figure 8. HPLC analysis of washed and unwashed cultures within a semi continuous model. Samples from neat (A) and washed (B) cultures taken for HPLC at the point of transfer to a fresh batch. g/L quantification was performed for glucose, whilst glycerol, ethanol and arabitol were quantified only with relative peak area. Data points represent the average of triplicate cultures.

The above results show that a semi continuous system which continuously refreshes the same cells into fresh media creates a bottleneck of non-oleaginous cells by the

third serial transfer. Conversely, it also showed that transferring cells once results in increased lipid production. An additional finding was that despite the increased lipid production of cells transferred once, little of the glucose was consumed. Clearly however, a process which refreshes the media just once is little more than an extended batch fermentation, with few of the added benefits of continuous or semi continuous bioprocessing.

To utilise this information within a semi continuous process, an experiment was conducted which mimicked the above, but introduced a low frequency of fresh cells from an SMB overnight culture every time cells were transferred to the next batch. In this set up, different strategies of culture removal/media addition were performed to simulate different semi continuous processes. For example, in a 1:5 process, 20% of the existing culture would be removed at each batch and topped up to volume with fresh media. In a 1:1 process, 50% of the culture would be removed and topped up with fresh media. In this high culture removal scenario it is expected that very few cells would go through three batches due to the amount of culture removed at each step, allowing the majority of cells to only ever reach the high lipid production stage of the second batch. At each point where culture was removed and media added, 1 mL of overnight culture was centrifuged, resuspended in 100 mL PSB, and added to the main culture.

This experiment began with cells from batch six of the previous experiment (Figure 7), which had become dead-ended with non-oleaginous cells. This culture was diluted at the 1:1, 1:3 or 1:5 ratios as described, with fresh cells from 1 mL of SMB overnight added to each. As expected, the 1:1 diluted culture had the highest average cell size and 1:5 the lowest (Figure 9). Comparing the respective histograms show the differences between the two populations in more detail, with it clear that the smaller cells from the previous culture at a lower frequency in the 1:1 system due to the higher culture removal (Figure 10). This trend continues within the next batch, with the average cell size for all ratios increasing further and the difference between the three strategies widening. Although further batches would need to be performed to evaluate long term longevity, the expected trend that smaller cells were gradually

being cultured out appears to be occurring, particularly for the 1:1 ratio semi continuous process. Although adding fresh cells to an existing culture is not typical within semi continuous cultures, the volume of cells added at each point as a fraction of the total number of cells was very low, and therefore could be a conceivable process when performed at scale.

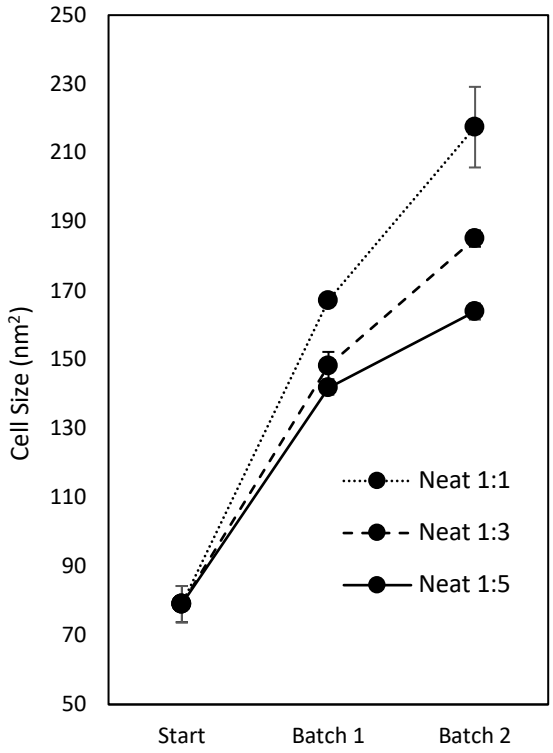


Figure 9. Model semi continuous culturing with varying ratios of culture addition/removal and inoculation of fresh cells. Cells from figure 7 batch six (washed) were pooled and used as the 'start' culture. This culture was mixed with fresh NLB media according to the ratios stated, i.e. in the 1:1 ratio, 5 mL of the starting culture was mixed with 5 mL fresh NLB, in the 1:5 ratio, 8 mL of the starting culture was mixed with 2 mL fresh NLB. At each point the cultures were mixed, 1 mL of washed overnight culture was inoculated into the culture.

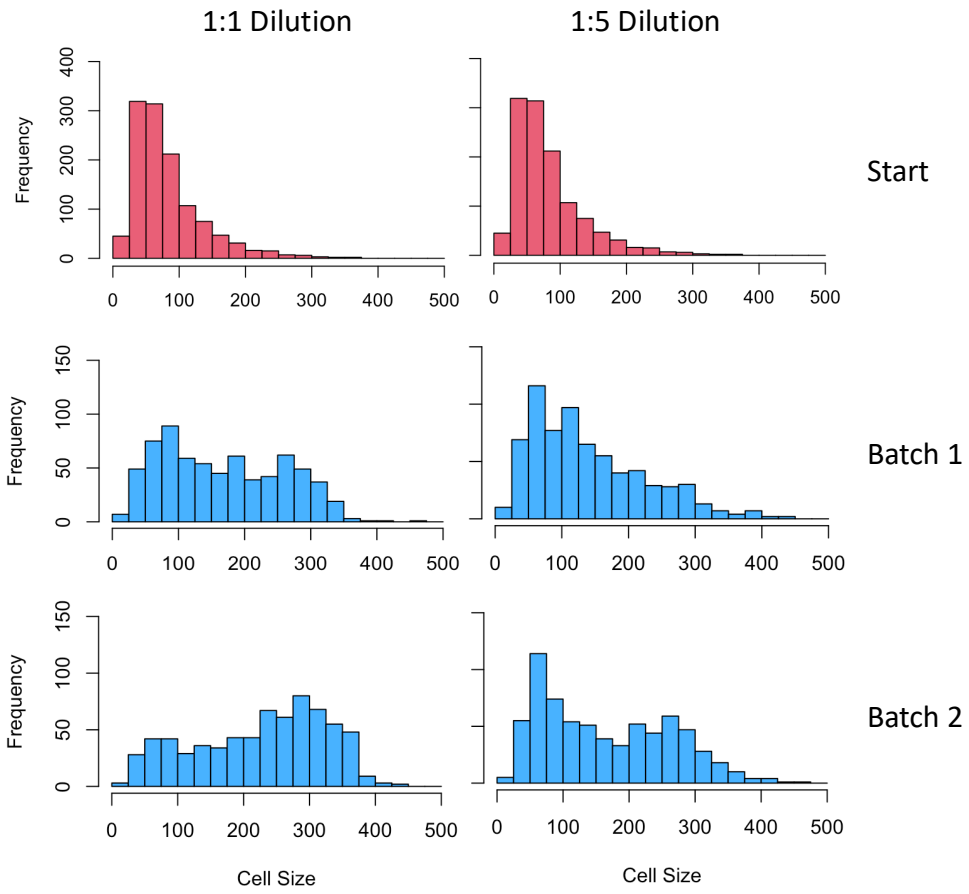


Figure 10. Cell population dynamics within model semi continuous culturing with varying ratios of culture addition/removal and inoculation of fresh cells. Cells from figure 7 batch six were pooled and used at the 'start' culture. This culture was mixed with fresh NLB media according to the ratios stated, i.e. in the 1:1 ratio, 5 mL of the starting culture was mixed with 5 mL fresh NLB, in the 1:5 ratio, 8 mL of the starting culture was mixed with 2 mL fresh NLB. At each point the cultures were mixed, 1 mL of washed overnight culture was inoculated into the culture. Histograms represent data from 900 cells, with 300 cells taken from each triplicate culture.

3.4.3. Scale up test of semi continuous model

A 250 L bioreactor was performed to assess this semi continuous model, although due to running restrictions the model had to be altered slightly. In this adapted model, the initial culture was inoculated and grown for three days at which point 50% of the culture was removed and topped up with fresh media and left to grow over the weekend. In the week following, 25% of the culture was removed and topped up every day until the Friday, where the 50% removal again occurred. The culture was finished according to the 25% daily removal schedule. At each point

where media was added/removed, 2x 1 L SMB overnight culture was also added to provide new cells to the bioreactor. Results from this 16 day experiment give a promising indication of this semi continuous model, as both dry cell weight and lipid production from day six onwards maintains a near-consistent level (Figure 11). It is particularly encouraging that when 50% of the culture is removed and replaced with fresh media before the weekend, particularly on the second occasion, that lipid production is not affected and biomass returns to a peak. The slight reduction in biomass and lipid accumulation by day 16 could be attributed to only removing 25% of the culture at a time, allowing small cells to accumulate before they are removed by the 50% media changeover over the weekend.

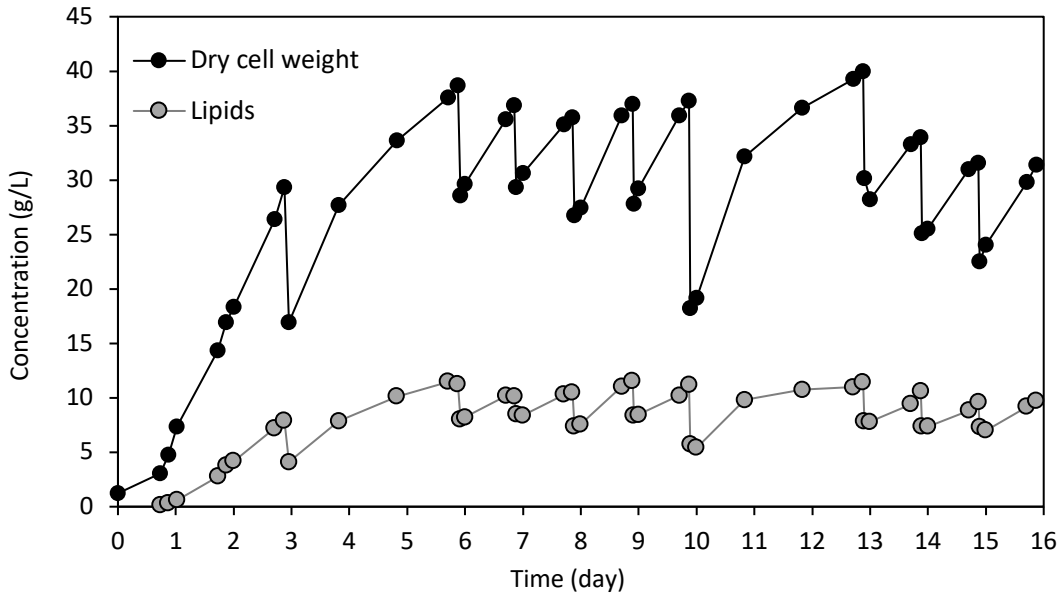


Figure 11. 250 L semi-continuous culture. 250 L of simplified NLB was prepared, and inoculated with 2 L of an SMB overnight culture. The culture was grown until day 3, at which point 50% of the culture was removed and topped up to volume with fresh NLB. On days 6, 7, 8, 9, 13, 14 and 15 25% of the culture was removed and topped up to volume. On day 10, 50% of the culture was removed and topped up to volume. At each point where culture was removed and topped up with fresh media, a 1 L SMB overnight was added at the same time.

3.5. Conclusion

The results presented in this study evaluate the factors affecting lipid production as a function of the environmental conditions likely to be relevant within semi

continuous bioprocessing. From this, two main conclusions were found. Firstly, it was shown that the lipid production of new *M. pulcherrima* cells can be negatively affected by the presence of the supernatant from an established culture. This effect appeared most significant when the supernatant contained a higher concentration of ethanol, though it is also possible that causative compounds or proteins may not have been identified within the analytical method used within this study. Secondly, it was determined that a lipid production bottle-neck occurred when *M. pulcherrima* cells were continuously cultured, with almost the entire population of cells displaying a non-oleaginous phenotype by the sixth transfer. Interestingly, cells transferred to fresh media once showed the highest lipid accumulation before production crashed when transferred further. To overcome this, a semi continuous model was developed whereby fresh cells were supplemented into the main culture at a low frequency to coincide with the point of culture removal and media addition. By adopting a process which removed 50% of the culture and replaced this volume with new media and a low dose of fresh cells, the production plateau was overcome as the bulk of cells were unlikely to remain within the culture for longer than two batches. The findings within this study are particularly important to the production of a microbial oil, which due to the volume of product required to be an economically viable process, will rely on continuous or semi continuous culturing to reduce costs and maintain consistent yields.

Acknowledgements

Thanks go to Felix Abeln, Mauro Moreno and Hadiza Auta for our collaborative work at the BEACON bioprocessing facility in Aberystwyth. Thanks also to those at the BEACON facility for their assistance during our time there. This research has been funded by the Industrial Biotechnology Catalyst (Innovate UK, BBSRC, EPSRC) to support the translation, development and commercialisation of innovative Industrial Biotechnology processes (EP/N013522/1) and by the EPSRC through the Centre for Doctoral Training in Sustainable Chemical Technologies (EP/L016354/1).

3.7. References

1. Balan, V. Current Challenges in Commercially Producing Biofuels from Lignocellulosic Biomass. *ISRN Biotechnol.* 2014.
2. Lee, J.; Lee, S.Y.; Park, S.; Middelberg, A.P.J. Control of fed-batch fermentations. *Biotechnol. Adv.* 1999.
3. Brethauer, S.; Wyman, C.E. Review: Continuous hydrolysis and fermentation for cellulosic ethanol production. *Bioresour. Technol.* 2010.
4. Stark, D.; Münch, T.; Sonnleitner, B.; Marison, I.W.; Von Stockar, U. Extractive bioconversion of 2-phenylethanol from L-phenylalanine by Saccharomyces cerevisiae. *Biotechnol. Prog.* 2002, *18*, 514–523.
5. Bauer, R.; Katsikis, N.; Varga, S.; Hekmat, D. Study of the inhibitory effect of the product dihydroxyacetone on Gluconobacter oxydans in a semi-continuous two-stage repeated-fed-batch process. *Bioprocess Biosyst. Eng.* 2005.
6. Hicks, R.H.; Sze, Y.; Chuck, C.J.; Henk, D.A. Enhanced inhibitor tolerance and increased lipid productivity through adaptive laboratory evolution in the oleaginous yeast Metshnikowia pulcherrima. Manuscript submitted for publication. 2018.
7. Hicks, R.H.; Chuck, C.J.; Scott, R.J.; Leak, D.J.; Henk, D.A. Comparison of Nile red and cell size analysis for high throughput lipid estimation within oleaginous yeast. Manuscript submitted for publication. 2018.
8. Ghareib, M.; Youssef, K.A.; Khalil, A.A. Ethanol tolerance of Saccharomyces cerevisiae and its relationship to lipid content and composition. *Folia Microbiol. (Praha).* 1988.

This declaration concerns the article entitled:													
Optimisation of semi continuous bioprocessing through cell dynamics monitoring.													
Publication status (tick one)													
draft manuscript	**X**	**Submitted**		**In review**		**Accepted**		**Published**					
Publication details (reference)													
Candidate's contribution to the paper (detailed, and also given as a percentage).	The candidate contributed to/ considerably contributed to/predominantly executed the… Formulation of ideas: This work was conceptualised by myself after initial results obtained in the previous chapter. 100% Design of methodology: Methodology was designed in its entirety by myself. 100% Experimental work: All work at small scale was performed by myself. Felix Abeln ran the samples HPLC samples. 90% The 250 L bioreactor was a collaborative project between BEACON at Aberystwyth University and our research group at the University of Bath. Felix Abeln, Mauro Moreno and Hadiza Auta and myself were involved on the Bath University side. Felix Abeln planned and designed the culturing schedule for the experiment. Presentation of data in journal format: Formatted for: European Journal of Lipid Science and Technology.												
Statement from Candidate	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.												
Signed								**Date**					

4. Enhanced inhibitor tolerance and increased lipid productivity through adaptive laboratory evolution in the oleaginous yeast *Metshnikowia pulcherrima.*

4.1. Commentary.

This chapter, and those following, detail the second half of this thesis where adaptive evolution was used to develop or improve industrially relevant phenotypes within *M. pulcherrima*, and the subsequent genomic analysis to find causative mutations. Unfortunately, as an entirely new method of lipid quantification was required rather than adapting an existing published fluorescence method, the cell size analysis technique doesn't feature within the analysis of the evolved strains.

The selection of the specific inhibitors used was based off data obtained from our collaborators in York University who provide hydrolysed material for our group in Bath University. Concentrations of inhibitors selected are justified within the text. Concentration of carbon source used is again based upon typical values found within hydrolysed material from York.

Enhanced inhibitor tolerance and increased lipid productivity through adaptive laboratory evolution in the oleaginous yeast *Metshnikowia pulcherrima.*

Robert H. Hicks,[a] Yuxin Sze,[b] Christopher J. Chuck,[c] Daniel A. Henk.[b]*

[a] Centre for Doctoral Training in Sustainable Chemical Technologies, Department of Biology and Biochemistry, University of Bath, Bath, BA2 7AY, United Kingdom

[b] Department of Biology and Biochemistry, University of Bath, Bath, BA2 7AY, United Kingdom

[c] Department of Chemical Engineering, University of Bath, Bath, BA2 7AY, United Kingdom

Abstract

Microbial lipid production from second generation feedstocks presents a sustainable route to future fuels, foods and bulk chemicals. The oleaginous yeast *Metshnikowia pulcherrima* has previously been investigated as a potential platform organism for lipid production due to its ability to be grown in non-sterile conditions and metabolising a wide range of oligo- and monosaccharide carbon sources within lignocellulosic hydrolysates. However, the generation of inhibitors from thermochemical pretreatment of these feedstocks causes downstream bioprocessing complications, and despite *M. pulcherrima's* comparative tolerance, their presence is deleterious to both biomass and lipid formation. Using either a single inhibitor (formic acid) or an inhibitor cocktail (formic acid, acetic acid, fufural and HMF), two strategies of adaptive laboratory evolution were performed to improve *M. pulcherrima's* fermentation inhibitor tolerance. Using a sequential batch culturing approach, the resulting strains from both strategies had increased growth rates and reduced lag times under inhibiting conditions versus the progenitor. Interestingly, the lipid production of the inhibitor cocktail evolved strains markedly increased, with one strain producing 41% lipid by dry weight compared to 22% of the progenitor. The evolved species was cultured in a non-sterile 2L stirred tank bioreactor and accumulated lipid rapidly, yielding 6.1 g/L of lipid (35% cell dry weight) within 48 hours; a lipid productivity of 0.128 g $L^{-1}h^{-1}$. Furthermore, the lipid profile was analogous to palm oil, consisting of 39% C16:0 and 56% C18:1 after 48 hours.

Key words:

Adaptive laboratory evolution, Microbial Oil, Lignocellulose, Fermentation Inhibitors, Bioprocessing, Biodiesel

4.2. Introduction

Microbial lipids, produced from heterotrophic organisms, are a versatile chemical feedstock that can be used to produce alternatives to fossil derived fuels and chemicals. In comparison to higher plant oils, such as palm oil, microbial oils can be produced on non-arable land, that does not compete with virgin rainforest or food production. While the lipid profile from microalgae is highly variable, generally, oleaginous yeast produce lipid profiles akin to plant oils with elevated levels of oleic and palmitic acid [1]. Recently, we reported on the oleaginous yeast *Metschnikowia pulcherrima* that can be grown in non-sterile conditions, while having the ability to metabolise a range of oligosaccharide and monosaccharide carbon sources [2–4]. Despite this potential, for microbial lipids to financially compete with both fossil and vegetable oils, a low cost lignocellulosic feedstock must be used as the feedstock source.

Processing lignocellulosic feedstocks presents multiple challenges, however, one key issue is the formation of inhibitory by-products resulting from the degradation of monosaccharides produced from the cellulose and hemicellulose feedstocks. These inhibitors halt or inhibit growth and metabolism, slow down productivity and reduce overall product yields [5]. Fermentation inhibitors can be classified according to their functional group; carboxylic acid, ketone, phenolic or aldehyde [6]. Two of the most abundant are 5-hydroxymethyl furfural (HMF) and 2-furaldehyde (furfural), resulting from the dehydration of hexose or pentose sugars respectively. The most common inhibitory acids produced are formic and acetic acid, with formic acid produced as a derivative of furfural or HMF, and acetic acid produced when acetyl groups are released from hemicellulose [7].

The inhibitory mechanism for these components is complex. HMF and furfural are non-specifically reactive with RNA, DNA and proteins as well as causing membrane damage resulting in disruptions to metabolism and cell viability [8]. Acetic and formic acid can enter yeast cells via diffusion through the plasma membrane where they dissociate into acetate/formate and a proton once within the cytoplasm.

Accumulation of protons leads to cytoplasm acidification causing metabolism impairment by inhibiting glycolytic enzymes and NADH dehydrogenases to increase lag times and reduce growth [7]. In addition to the individual effects of each inhibitor, they act synergistically to create negative effects that exceed the sum of each individual effect [9,10].

A rational engineering approach to improve inhibitor tolerance is complex due to the vast network of molecular mechanisms involved. In addition, as genetic toolkits to accomplish such aims are unavailable within many non-model organisms, alternative phenotypic improvement strategies, such as adaptive laboratory evolution (ALE), are potentially more promising. ALE is a method whereby microorganisms are continually cultured under a selective pressure in defined conditions for periods of time to allow for the selection of advantageous phenotypes [11]. Whilst ALE often begins with a natural isolate, it is not uncommon to first perform random mutagenesis through chemical or UV means[12]. Many different evolution strategies have been applied to improve the biotechnological capacity of natural or genetically engineered *Saccharomyces cerevisiae* strains, for example improving tolerance to ethanol [13] or lignocellulosic fermentation inhibitors [14] as well as improving the fermentation of non-preferred sugars such as xylose [15] and arabinose [16]. There are limited examples of strain improvement through ALE within oleaginous yeast, though *Rhodococcus opacus* strains which had first undergone genetic engineering followed by mutagenesis selection were further improved through sequentially batch culturing within increasing concentrations of lignin [17].

Oleaginous yeast such as *Rhodotorula glutinis*, *Yarrowia lipolytica* and *Lipomyces starkeyi* tend to have a higher inhibitory tolerance compared to *S. cerevisiae* [1,2]. Similarly, *Metschnikowia pulcherrima* displays a naturally high tolerance to furfural, acetic acid and HMF [3]. *M. pulcherrima*'s industrial potential is further improved by its inherent antimicrobial activity, allowing low cost non-sterile cultures to be performed [18].

The aim of this current study was to increase the fermentation inhibitor tolerance of *M. pulcherrima* whilst comparing two ALE strategies; either using a fermentation inhibitor cocktail (containing furfural, HMF, formic acid and acetic acid) as a selective pressure, or formic acid in isolation.

4.3. Methods.

4.3.1. Chemicals.

Unless otherwise stated, chemicals were sourced from Sigma Aldrich and used without further purification.

4.3.2. Strains, strain maintenance and media.

The *M. pulcherrima* strain (NCYC2580) used as the progenitor in this study was obtained from the National Collection of Yeast Cultures. Strains were maintained on malt extract agar (MEA) plates, and re-streaked on a fortnightly basis. For the preparation of overnight cultures, a single colony was inoculated into 10 mL SMB pH 5 (3% tryptic soy broth, 2.5% malt extract), and incubated at 25 °C with 200 rpm agitation. Optical densities throughout were measured at 595_{nm}. Media used within ALE study was a nitrogen limited broth (NLB) consisting of: Glucose 40 g/L, $(NH_4)_2SO_2$ 2 g/L, KH_2PO_4 7 g/L, $MgSO_4$ $7H_2O$ 1.5 g/L, $NaHPO_2$ 2 g/L and yeast extract 1 g/L. This media was autoclaved without glucose or $MgSO_4$, which were added separately after autoclaving individual stock solutions. YNB medium was prepared as follows - 1.78 g/L Yeast Nitrogen Base w/o amino acids, 25 g/L glucose and variable amounts of $(NH_4)2SO_2$ per desired concentration.

4.3.3. Evolution experiment.

Three growth conditions were used for the initial screening experiment: NLB, NLB + 0.6 g/L formic acid and NLB + inhibitor cocktail (0.7 g/L of furfural and acetic acid, and 0.35 g/L formic acid and HMF). Triplicate overnight cultures grown in SMB were

diluted to an OD ~ 1 with PBS the following morning. Each diluted culture was used to inoculate each of the three treatment broths by adding 500 µL into 10 mL media within a glass culture tube. Cultures were incubated at 25 °C with 200 rpm agitation with OD measurements performed at regular intervals through the 72 hour period.

For the ALE experiment, cultures were started as above using the same starting formic acid and inhibitor cocktail concentrations, with five replicate lineages for each condition. OD measurements of each culture were taken at 24 and 48 h intervals, with cultures transferred after 48 hours by inoculating 100 µL of each culture into a fresh 10 mL culture tube. A culture sample was taken for storage at each transfer point. In the event of no growth in one or more of the culture tubes after 48 hours, successful lineages were expanded to seed new tubes to replace those lost, maintaining a total of five culture tubes per experiment. Inhibitor concentrations were increased when batches were reaching stationary phase after 48 hours to ensure that as much as practically possible exponentially growing cells were being transferred to the next batch. The evolution experiment was concluded after approximately 1000 hours where the final strains were glycerol stocked, and plated onto MEA to provide single colonies for phenotypic analysis.

4.3.4. 96 well plate growth phenotypic assays.

To phenotypically analyse the resulting mutant cell lines, a 96 well plate format was used. To perform these, triplicate overnight cultures for each mutant strain, as well as the progenitor were prepared as stated previously, with each diluted to an OD ~ 1 the following morning. To prepare the 96 well plate, 140 µL of media was added to the appropriate well, and to this, 10µL of each diluted culture was added. Media blanks were plated as negative controls. Cultures were incubated within a shaking plate reader with temperature maintained at 25 °C. Optical density was read at 30 minute intervals, for a period of 48 hours unless otherwise stated within figure legends. Max growth rates were calculated using RStudio [19], using Ln(OD) values. Media composition and inhibitor concentrations used within the phenotypic screening assays are described within figure legends.

4.3.5. Lipid extraction and profiling.

Overnight cultures were prepared as previously stated, with 500 µL of the diluted overnight culture used to inoculate 10 mL of NLB or NLB supplemented with an inhibitor cocktail (concentrations used for this assay were: Acetic acid and furfural 0.7 g/L, formic acid and HMF 0.35 g/L.
The lipid extraction methodology is based upon that proposed by Bligh and Dyer [20], and modified to the following:

Upon completion of culturing, 9 mL of cells were centrifuged at 11.3 thousand RCF, resuspended in 1 mL PBS, and centrifuged again, discarding the supernatant and freezing cell pellets instantly in liquid nitrogen. Following this, frozen cell pellets were freeze dried at -40°C for a minimum of six hours. For the cell disruption, a preweighed amount of cell pellet (ideally within the range of 10 – 100 mg) was mixed with 10 mL 6M HCl, and stirred at 80 °C for one hour. To extract lipids, 10 mL of chloroform:methanol (1:1) was added, and the mixture was stirred overnight at room temperature. To quantify the lipid weight, the lower chloroform phase was removed by hand with a glass pipette, avoiding the emulsion layer which forms between the chloroform and aqueous phase. The chloroform was then fully evaporated via rotary evaporator at 50 °C, and the remaining lipids were weighed to determine the lipid weight % of the cells.

For the lipid fatty acid profile, extracted lipid samples were transesterified with methanol and an additional 1% H_2SO_4, heated at 90 °C under pressure for three hours. The resulting fatty acid methyl esters were extracted with hexane, which was washed with water to remove any residual glycerol or sulfuric acid. GC-MS analysis was carried out using the Agilent 7890A Gas Chromatograph equipped with a CP-Sil capillary column (25 m x 0.250 mm internal diameter) and a He mobile phase (flow rate: 1.2 ml min^{-1}), coupled with an Agilent 5975C MSD. Approximately 50 mg of each sample was dissolved in 100 ml hexane and 1 µl of each solution was loaded onto the column, pre-heated to 40 °C. This temperature was held for 1 minute and

then heated to 250 °C at a rate of 10 °C min^{-1} and then held for 10 minutes. The FAME profile was calculated in reference to known standards.

4.3.6. Bioreactor culturing.

Overnight cultures were inoculated using a 30 mL SMB culture with an OD between 7-10 which was centrifuged and resuspended in 10 mL PBS. Culturing volume used was 1L, with 5 mL polypropylene glycol P2000 (antifoam) added upon inoculation. 80% oxygenation was maintained at the determined concentration through agitation (900 rpm max) and sparging (max 3:1 v/v), pH was maintained at 4 through the addition of nitric acid (1M) or sodium hydroxide (1M) and temperature was maintained at 20 °C. These parameters were controlled by the Fermac 320 bioreactor control unit. The cultures were performed to replicate low-cost industrial conditions, and as such, were under non-sterile conditions. Media used was NLB, or NLB supplemented with 0.7 g/L of furfural and acetic acid, and 0.35 g/L formic acid and HMF.

4.4. Results.

4.4.1. NCYC2580 inhibitor tolerance and adaptive laboratory evolution.

Concentrations of 0.6 g/L formic acid for the single inhibitor strategy and an inhibitor cocktail containing 0.7 g/L of furfural and acetic acid + 0.35 g/L formic acid and HMF for the multi inhibitor investigation were found to reduce 48h growth by approximately 50% compared to NLB without inhibitors, and therefore chosen as the starting concentrations for the ALE experiments (Figure 1). Inhibitor cocktail cultures displayed an extended lag time compared with the other two conditions, failing to reach an OD greater than 1 until approximately 29 hours. Despite not causing an increased lag time, the formic acid supplemented cultures instead appear exhibit biphasic growth.

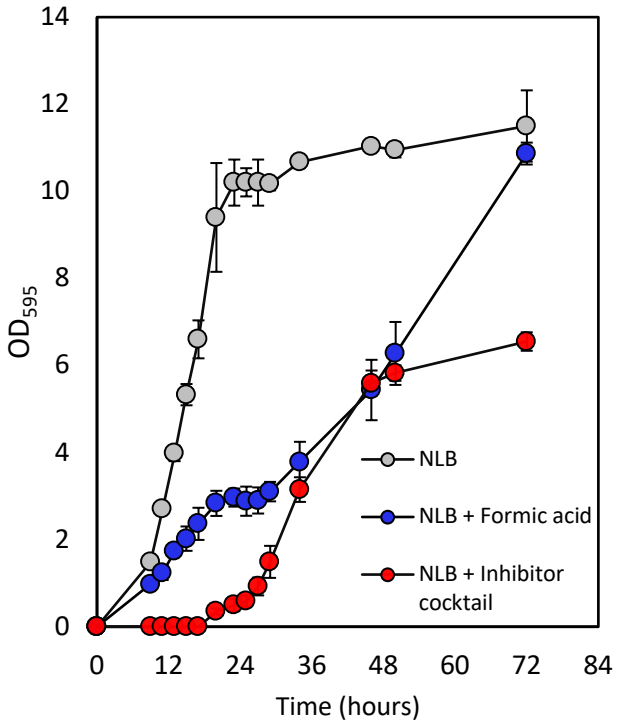


Fig. 1. Growth profiles of NCYC2580 in the presence of inhibitors. Strains were grown in NLB as a control (grey circles), NLB with 0.6 g/L formic acid (blue circles) and NLB with an inhibitor cocktail (0.7 g/L furfural and acetic acid, and 0.35 g/L formic acid and HMF; red circles). Error bars represent the mean and standard deviation of triplicate cultures

For both ALE experiments, five clonal lineages were started in parallel, and in the event of lineage loss, a successful culture was expanded to maintain a total of five cultures. Cultures were transferred after 48 h, with 100 µL of undiluted culture used to inoculate the following batch. Inhibitor concentrations were increased when the 48 hour OD indicated stationary phase was being reached, as determined by the control growth curve in Figure 1. Using this approach, the majority of cells transferred to the new batch would be growing exponentially.

ALE: Formic acid.

With a starting concentration of 0.6 g/L formic acid, OD was consistently high and the concentration across all five lineages was increased to 0.7 g/L at batch 4 (Figure 2). Increasing the formic acid concentration to 1 g/L at batch 10 resulted in reduced growth, characterised by an increased lag time (shown by the 24 h sample) in batches

10 and 11, and a decrease in the transfer OD through batches 12 to 14. All five cultures stabilised from batch 15 onwards, allowing the concentration to be increased further. Though the initial batch at 1.2 g/L formic acid saw reduced growth across all lineages, the following four batches recovered, indicating adaptation to this concentration. In total, 22 batches were performed within which a doubling of the initial formic acid concentration of 0.6 g/L to 1.2 g/L was achieved.

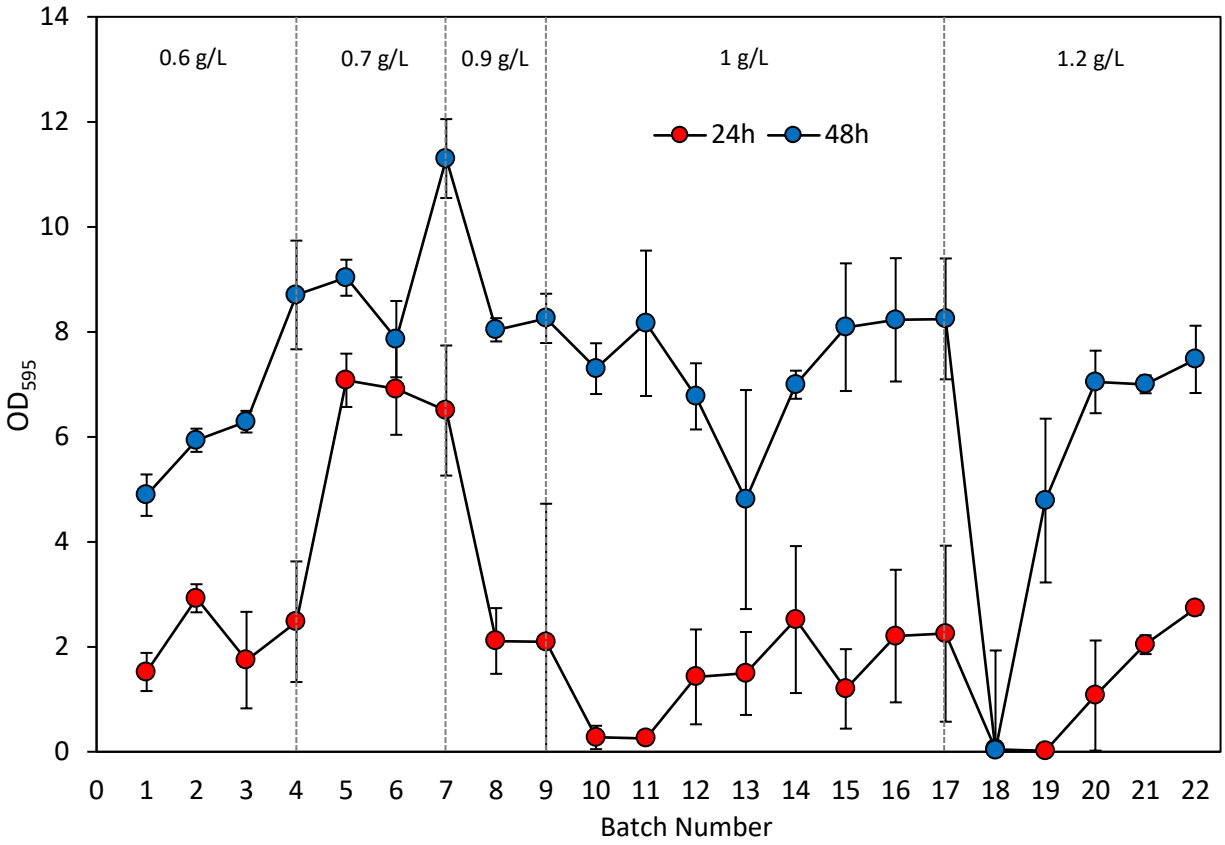


Fig. 2. Overview of the NCYC2580 formic acid adaptive laboratory evolution experiment. Five colonies were used to inoculate five individual NLB + 0.6 g/L formic acid cultures. Cultures were transferred to fresh media after 48h growth; red and blue circles represent the OD values taken at 24 and 48 hours respectively for each batch. All five lineages continued throughout the duration of experiment, therefore error bars represent the mean and standard deviation of five cultures. The concentration of formic acid present in each batch as the experiment progresses is indicated within the plot

ALE: Inhibitor cocktail

The second ALE strategy applied a cocktail of four inhibitors as the selective pressure, which unlike the formic acid experiment, saw frequent lineage loss within the first 9 batches (Figure 3a). In total, only one of the initial five lineages completed the entire ALE experiment, itself frequently used to expand out to five cultures. The OD data correlates with this batch to batch instability, with large fluctuations in both the 24 and 48 hour OD sample for the first 11 batches (Figure 3b). The trend onwards suggests that adaptation to these conditions; for instance, increases to 24 hour OD measurements indicated that advantageous adaptations were occurring to favour reduced lag time. These results were in correlation with those in Figure 3a.

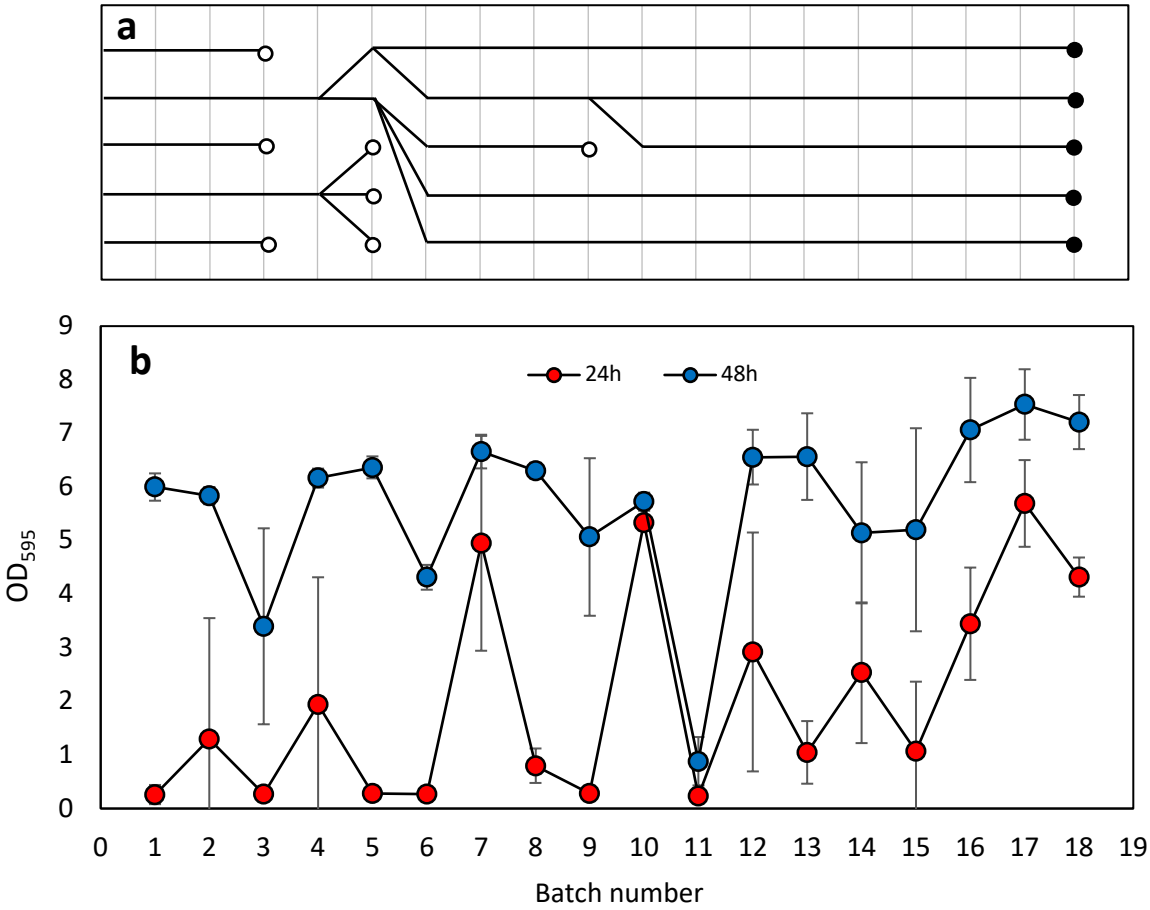


Fig. 3. Overview of the NCYC2580 inhibitor cocktail adaptive laboratory evolution experiment. Five individual colonies were used to inoculate five NLB + inhibitor cocktail (0.7 g/L of furfural and acetic acid, and 0.35 g/L formic acid and HMF) cultures. a - ALE lineage tree showing successful batches replacing failed batches (clear circles) to maintain five cultures. b - Cultures were transferred to fresh media after 48h; red and blue circles represent the OD values taken at 24 and 48 hours respectively for each batch. Error bars represent the mean and standard deviation of five cultures

4.4.2. Phenotypic Analysis of Evolved Strains

Upon completion of both ALE experiments, a streak plate was made from each final batch culture onto MEA agar, and a single colony from each plate was isolated onto a new MEA plate. Four out of the five lineages from each evolution strategy were taken forward for phenotypic analysis and compared to the progenitor strain. They were named as follows:

Formic acid evolved strains: F1, F2, F3, F4.
Inhibitor cocktail evolved strains: 4x1, 4x2, 4x3, 4x4.

Maximum growth rates in both rich (SMB) and defined (YNB + glucose) media determined that the evolved strains performed comparably to the progenitor when assayed in media different from the evolutionary experiments, suggesting that no discernible evolutionary trade off was observed when cells were grown in these two medias had occurred (Figure 4). Strains were assayed in NLB supplemented with an inhibitor cocktail of 1 g/L acetic acid and furfural, and 0.5 g/L formic acid and HMF. In this medium, differences in the growth rate were observed, with strains evolved to the inhibitor cocktail displaying maximum growth rates double that shown by the progenitor strain (Figure 5a). The strains evolved within formic acid supplemented media also showed increased growth rates versus the progenitor, though to a lesser extent. Only the formic acid evolved strains showed increased growth rates when assayed in NLB + 0.85 g/L formic acid (Figure 5c) meaning that despite both evolution strategies yielding strains with increased growth rates when assayed with an inhibitor cocktail, here, only the formic acid evolved strains showed an increased growth rate. A decrease in the lag time across all evolved strains versus the progenitor was observed in both conditions (Figure 5b and 5d). This was most notable when grown in NLB + inhibitor cocktail where the lag time reduced by over half, with growth occurring as early as 13 hours post inoculation versus around 40 hours for the progenitor.

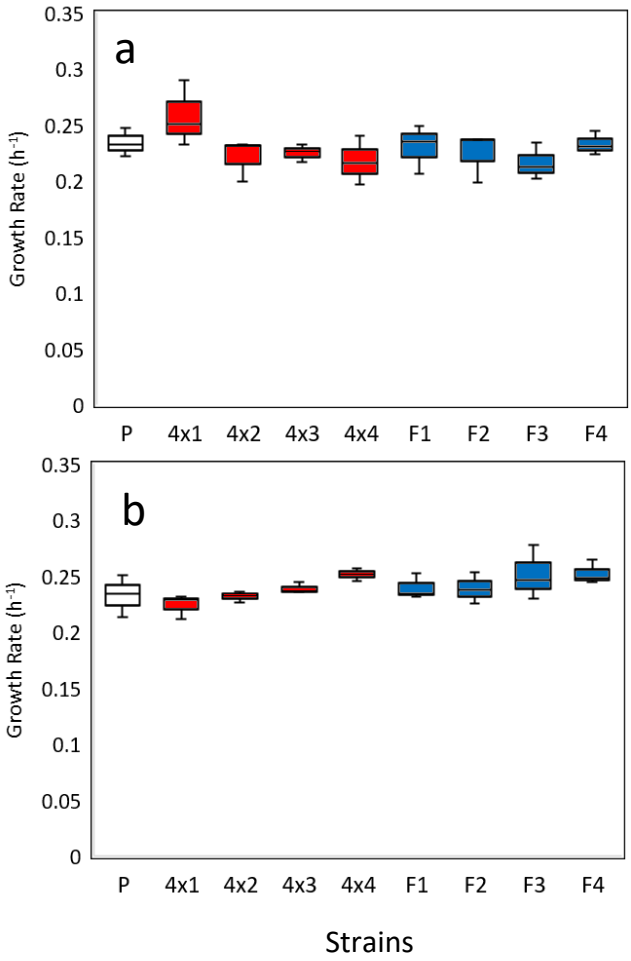


Fig. 4. Maximum growth rates of eight mutant NYC2580 strains and the progenitor in **(a)** SMB and **(b)** YNB + glucose. Each strain was assayed in triplicate with growth performed within a 96 well plate. Max growth rates were calculated using R Studio. Red – inhibitor cocktail strains, blue – formic acid evolved strains

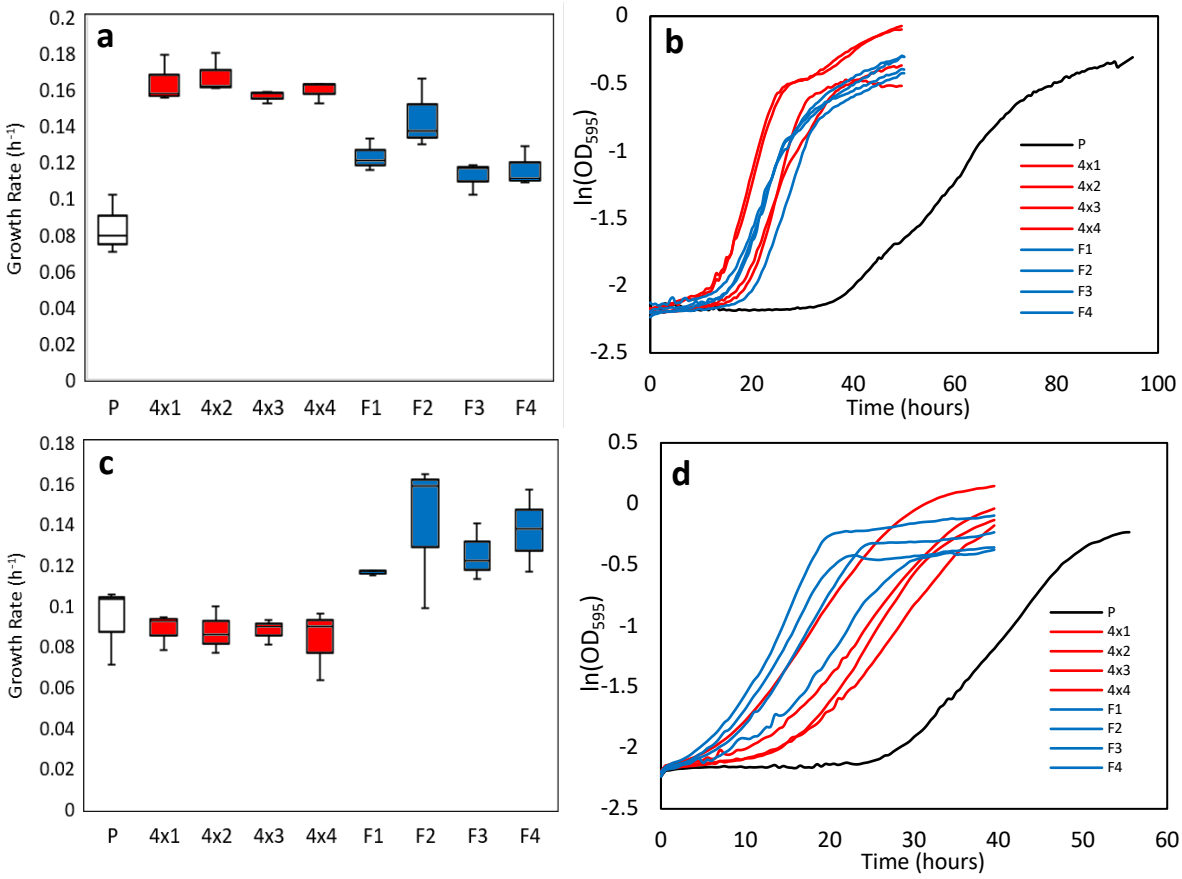


Fig. 5. Maximum growth rates **(a)** and plotted growth rates **(b)** of eight evolved NYC2580 strains and the progenitor in NLB + inhibitor cocktail. Inhibitor cocktail consisted of: acetic acid and furfural 1 g/L, formic acid and HMF 0.5 g/L**.** Maximum growth rates **(c)** and plotted growth rates **(d)** of eight mutant NYC2580 strains and the progenitor cultured in NLB + 0.85 g/L formic acid. Each strain was assayed in triplicate with growth performed within a 96 well plate. Max growth rates were calculated using R Studio, curves represent the ln(OD) of each growth curve at 30 minute intervals. Red – inhibitor cocktail strains, blue – formic acid evolved strains

4.4.3. Lipid production of evolved strains.

Given that evolutionary trade-offs can be a significant draw back to ALE strain improvement, it is possible that the selection of inhibitor tolerant phenotypes could result in the loss of the oleaginous phenotype. To test for any effect on the lipid production, seven day cultures under inhibiting and non-inhibiting conditions were performed for the evolved strains (Figure 6). Lipid production of the progenitor strain grown in NLB without inhibitors was slightly higher (22.1%) than the four lineages evolved in formic acid (12.3-17.4%). Lipid accumulation by the four-inhibitor cocktail

evolved strains increased against the progenitor to yield between 32.5 – 41% lipid by dry weight. Furthermore, despite the comparable biomass of the formic acid evolved strains and the progenitor, all inhibitor cocktail evolved strains achieved a higher biomass, with the best performing strain reaching an average of 14.4 g/L. This result was also observed when all strains were cultured within NLB + inhibitor cocktail, with the inhibitor cocktail evolved strains outperforming the others, including one mutant reaching an average biomass of 8.5 g/L with a lipid accumulation of 34.7%. The progenitor in this media had reduced biomass (4.3 g/L), but still achieved 20% lipid accumulation. In these conditions, however, the formic acid evolved strains had increased lipid production compared to when grown in control media (averaging 23% versus 15%), despite a lower biomass. Though formic acid and inhibitor cocktail evolved strains performed comparably within the growth assays, these results present strong evidence of unique adaptations occurring from each ALE strategy.

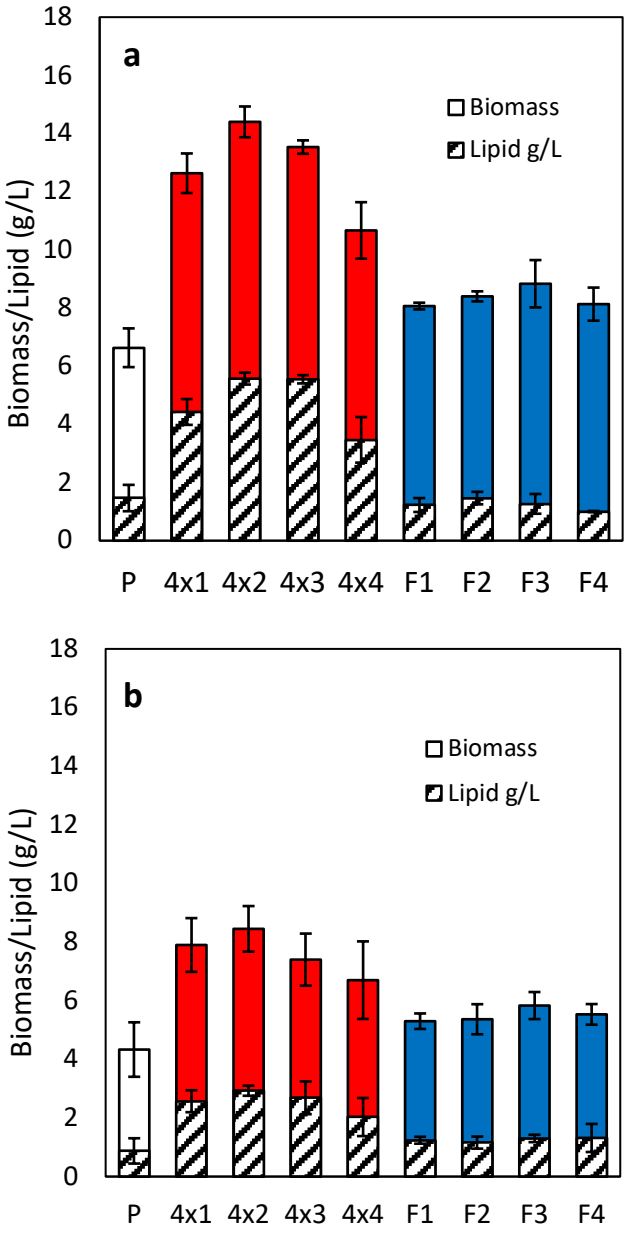


Fig. 6 . Lipid quantification and biomass of evolved strains versus the progenitor grown in NLB **(a)** and NLB + Inhibitor cocktail **(b).** Inhibitor cocktail contained acetic acid and furfural 0.7 g/L, formic acid and HMF 0.35 g/L. Triplicate samples were cultured for seven days for each strain. Error bars values represent the standard deviation of the mean

Using the same samples assayed for lipid production, FAME profiling of the lipid from all strains revealed further phenotypic differences. The FAME profile for the progenitor and inhibitor cocktail evolved strains varied very little when grown in the presence or absence of an inhibitor mix, though the latter had slightly increased ratios of C16:1 and C16:0 and lower amounts of C18:1 when compared to the

progenitor (Table 1). The formic acid evolved strains FAME profile varied considerably to the parental however, particularly when grown without inhibitors. Ratios of C16:0 decreased to an average of 10.6% across the four strains compared to 22.5% for the progenitor, and C18:1 and C18:2 levels increased to 78.5% and 8% on average compared to 71.3% and 2.5% respectively.

Table 1

FAME profiles of the progenitor and evolved strains.

NLB	C16:1	C16:0	C17:1	C18:2	C18:1	C18:0
Progenitor	3.1	22.5	-	2.5	71.3	0.5
4x1	4.4	25.6	-	1.2	67.3	1.5
4x2	8.5	25.9	0.3	1.6	61.4	2.3
4x3	7.9	25.3	0.3	2.5	62.0	2.0
4x4	3.4	24.2	-	1.4	68.9	2.1
F1	1.2	11.1	-	5.8	81.2	0.7
F2	2.4	12.3	-	7.2	77.2	0.8
F3	1.2	8.1	-	11.8	77.6	1.3
F4	3.2	11.0	-	7.2	78.2	0.5

NLB + Inhibitors	C16:1	C16:0	C17:1	C18:2	C18:1	C18:0
Progenitor	1.9	22.0	-	1.8	73.2	1.1
4x1	5.3	27.5	0.2	3.5	62.4	1.1
4x2	4.6	27.0	-	2.7	63.9	1.8
4x3	4.9	25.7	-	2.1	65.4	1.8
4x4	3.0	26.1	-	3.4	66.4	1.1
F1	1.9	17.6	-	1.7	74.9	3.9
F2	2.6	17.1	0.5	1.7	75.5	2.7
F3	3.0	16.1	0.7	1.0	76.7	2.5
F4	3.3	16.1	0.8	1.1	76.4	2.3

NLB + inhibitor cocktail (acetic acid and furfural 0.7 g/L, formic acid and HMF 0.35 g/L). Numbers represent fatty acid composition as % of total fatty acids

4.4.4. Influence of carbon-to-nitrogen ratio on lipid production in progenitor and evolved strains.

Two strains; 4x3 and F3, were taken forward with the progenitor for further investigation into carbon-to-nitrogen (C:N) response and how this has been affected by each directed evolution strategy. Lipid production was assayed within a full

factorial (2 levels) experimental design, with carbon and nitrogen as the input factors. The 2D surface plot models the lipid production by weight percent for each strain in response to varying C:N within the experimental space. The response range (Lipid %) for each plot ranges from the maximum and minimum lipid production values obtained for that individual strain, within the C:N ratios assayed (Figure 7). All strains exhibit a similar trend in response to varying C:N ratios, accumulating the minimum amount of oil when nitrogen is high and carbon is low and maximum amounts when carbon is high and nitrogen low. The response plot for 4x3 immediately differs from F3 and the progenitor in continuing to accumulate a high amount of oil when carbon concentrations drop from their highest point, as shown by the larger red area. Another striking difference between all strains is the lipid production when both carbon and nitrogen are low (bottom left; 10 g/L glucose, 0.3 g/L ammonium sulphate). Here, F3 accumulates a low amount of oil relative to its maximum, 4x3 accumulates oil closer to its maximum and the progenitor comes in between the two. Finally, further differences are seen between the strains when comparing lipid production in top right and bottom left corners of the experimental space (high carbon, high nitrogen and low carbon, low nitrogen respectively). Here, the progenitor correlates with F3, producing a similar level of lipid at both conditions, whilst production is high for 4x3 when carbon and nitrogen concentrations are low, and decreased when concentrations are higher.

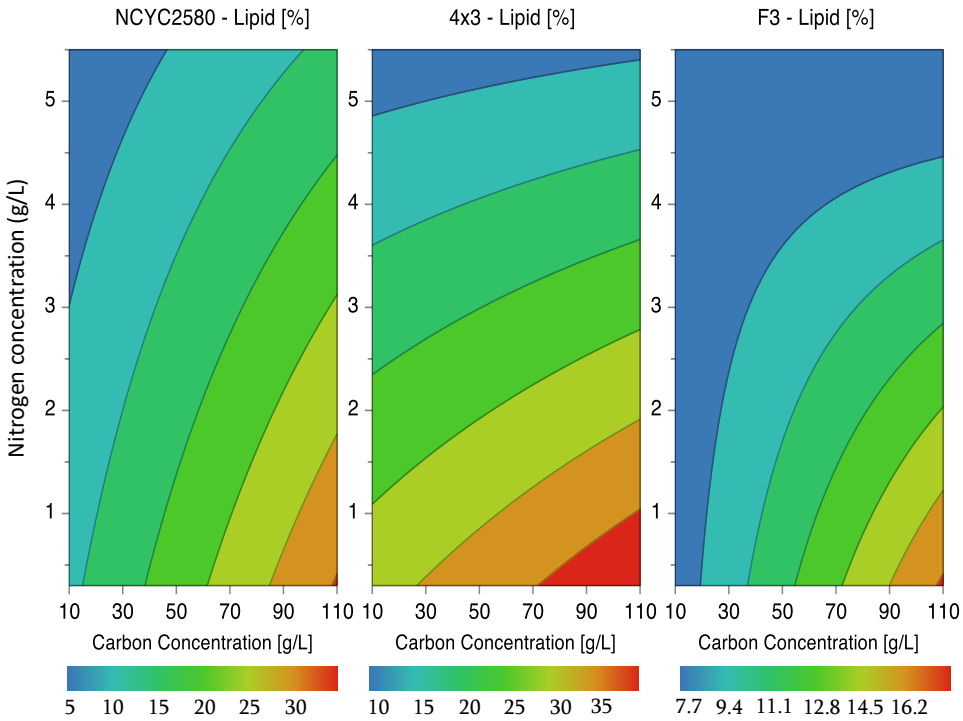


Fig. 7. 2D surface plot from a full factorial (2 levels) experimental design with the progenitor NCYC2580 and evolved strains F3 and 4x3 assayed for lipid production in response to changing concentrations of carbon and nitrogen within YNB media for seven days. Each plot represents the maximum and minimum lipid production in weight percent for the individual strain, with the range split into seven equal contours.

4.4.5. Culturing in 2L bioreactor under non-sterile conditions

To further assess the biotechnological potential of the evolved strains, the '4x3' strain evolved in the cocktail of inhibitors was taken forward for culturing in a 2L controlled bioreactor under inhibiting and non-inhibiting conditions. Here, both cultures were performed in a non-sterile manner to best represent an industrial process. Final biomass within both bioreactor conditions exceeded that observed on the smaller scale, with control media conditions achieving 16 g/L compared to 13.5 g/L within tubes, and 14.5 g/L versus 7.4 g/L with the inhibitor containing media. The biomass increase is likely due to the high oxygenation supplied within the bioreactor (80% DO) and better gas transfer. This increase in biomass also suggests that increased oxygenation aids inhibitor detoxification leading to overall higher growth.

Though final biomass at day seven within the control bioreactor was higher than within culture tubes, here, peak biomass of 17.5 g/L was achieved after just 48 hours, giving a biomass productivity of 0.356 g L^{-1} h^{-1}. A similar trend was also observed for lipid accumulation, where final values of 38.1% and 33.3% for control and inhibitor conditions respectively showed only a slight increase against 48 hour values of 35.1% and 29%. This equates to a lipid productivity of 0.128 g L^{-1} h^{-1} and 0.088 g L^{-1} h^{-1} respectively.

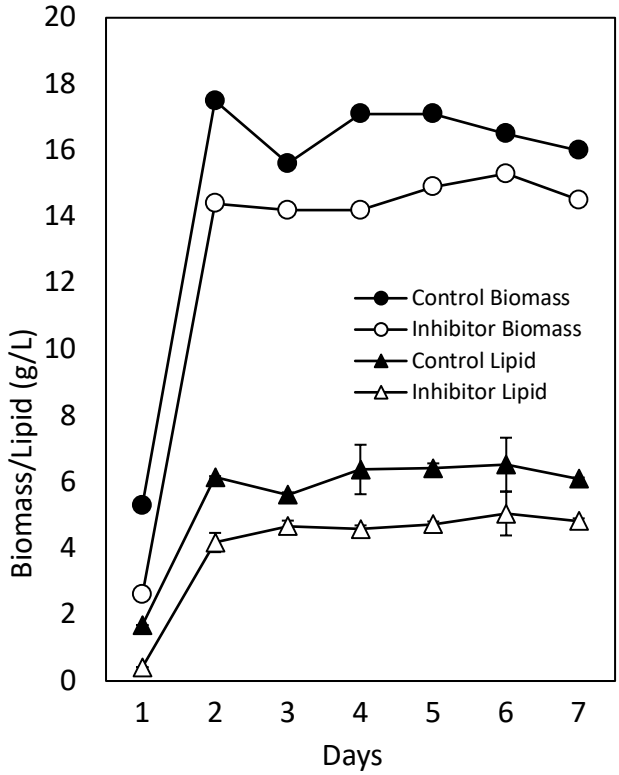


Fig. 8. Lipid quantification and biomass of strain '4x3' when grown in NLB (black circles/triangles) versus NLB + inhibitor cocktail (clear circles/triangles) under non-sterile bioreactor conditions. Inhibitor cocktail consisted of acetic acid and furfural 0.7 g/L, formic acid and HMF 0.35 g/L. Bioreactor was maintained at 20°C, pH 4 and 80% dissolved oxygen. Error bars for lipid production represent the standard deviation of triplicate samples taken at the same timepoint

Similarly to the smaller scale, the lipid profile did not change substantially when grown with or without inhibitors, however the level of saturated esters increased substantially compared with the tube cultures. For example, in the bioreactors the amount of C16:0 and C18:1 after 24 hours was 46% and 50% respectively against 25% and 62% when grown in the tubes (Table 2). This difference is presumably due to the increased oxygenation within the bioreactor.

Table 2. FAME profiles of the progenitor and evolved strains within bioreactors.

NLB	C16:1	C16:0	C18:2	C18:1	C18:0
Day 1	1.2	46.0	0.0	50.8	1.9
Day 2	3.2	38.8	1.3	55.8	0.9
Day 3	4.8	35.2	2.1	57.4	0.6
Day 4	3.9	37.3	0.0	57.6	1.2
Day 5	3.8	31.5	1.8	62.1	0.8
Day 6	3.8	30.3	1.9	62.9	1.1
Day 7	3.7	30.3	1.3	63.8	0.9

NLB + Inhibitors	C16:1	C16:0	C18:2	C18:1	C18:0
Day 1	0.0	54.9	0.0	45.1	0.0
Day 2	3.8	39.5	1.9	54.2	0.6
Day 3	4.0	34.5	1.7	59.0	0.8
Day 4	5.4	31.4	2.4	59.6	1.2
Day 5	5.8	30.0	2.4	60.6	1.1
Day 6	5.2	30.7	1.9	60.8	1.3
Day 7	3.8	30.6	1.6	63.3	0.8

NLB + inhibitor cocktail (acetic acid and furfural 0.7 g/L, formic acid and HMF 0.35 g/L). Numbers represent fatty acid composition as % of total fatty acids

4.5. Discussion

Adaptive Evolution of NCYC2580.

The 48 hour batch transfer experimental design used within this study aimed to maintain cells in exponential phase as much as practically possible. In this way, stationary phase adaptation was to be minimised, and cells were exposed to a 'fresh' dose of inhibitors when stress tolerance is generally at its lowest [21]. The short batch time was also designed to specifically select for two industrially advantageous phenotypes; high growth rate and low lag time under inhibiting conditions, with slower growing populations eventually dropping out due to successive low frequencies at transfer. This approach however, is not favourable for selecting survival adaptations once the stationary phase was reached.

Differences between both evolution strategies were stark; inhibitor concentrations in the formic acid strategy gradually increased to 1.2 g/L from a starting point of 0.6 g/L whilst the starting concentration within the inhibitor cocktail strategy was

maintained for the duration and lineages were frequently lost due to batch death. This result is most likely due to the multitude of stresses exerted by the synergistic effect of inhibitors [10]. Successful ALE experiments using multi-inhibitor containing lignocellulosic hydrolysate or synthetic media containing inhibitor cocktails using a sequential batch approach are limited within the literature. Rather, studies of this kind are performed within chemostats where the maintenance of biomass leads to quicker detoxification of inhibitors by reducing the dose per cell effect [10,22]. A sequential batch strategy applied by Koppram et al. in agreement with results within Figure 3**Error! Reference source not found.**, also encountered early complications, r eporting a lag time increase of 35 to 50 hours within the initial batches [23]. However, as batch transfers were performed upon depletion of glucose within the media opposed to at a set time point as presented here, it is likely that this strategy is not targeted at selecting for mutants with reduced lag time or increased growth rate unlike the rationale presented here.

Growth Analysis of Evolved Strains.

As growth rate improvements versus the progenitor were not observed within YNB + glucose and SMB (Figure 4), it is improbable that the ALE strategies have caused advantageous mutations or copy number amplifications that increase glycolytic efficiencies leading to decreased lag time and increased growth rate. Similarly, it is unlikely that the evolved strains are faster growing due to increased glucose transportation efficiencies as has been observed within *S. cerevisiae* strains evolved to glucose-limited media [24,25]. Rather, given that growth rate improvements were only observed under inhibiting conditions, then any adaptive improvements are likely to be specific to inhibitor tolerance. Mutation or amplification candidates aiding acid tolerance include multi-drug transporters or proteins within the ABC transporter family such as Pdr12 or Azr1, the latter of which has been shown to contribute to acetic acid tolerance [26,27]. Increased tolerance and growth within furfural and HMF containing medias on the other hand could implicate the efficiency or abundance of furfural/HMF reducing enzymes such as Adh1, Adh6 and Sfa1 [28]. Though there are limited examples of genomic sequencing applied to strains evolved

to fermentation inhibitors, Sako *et al*. found both a loss of heterozygosity within one chromosomal arm and duplication within another within an engineered *S. cerevisiae* strain evolved to inhibitor containing media, with both events involving members of the *PDR* ABC transporter gene family [29]. The loss of heterozygosity event occurred within the chromosomal arm containing *PDR1,* a gene regulating the expression of multidrug resistance transporters including those associated with tolerance to gallic acid, whereas the duplication event contained *PDR18*, a gene important for tolerance to herbicides and metal ions, *PDR16*, a regulator of lipid biosynthesis and *PDR17*, a regulator of membrane composition.

As two distinct evolution strategies were performed, it was anticipated that mutants will develop adaptations specific to their own evolution condition. In this respect, as the inhibitor cocktail contained both acid and aldehyde inhibitors, it was expected that these mutants would perform well when assayed in just formic acid supplemented media which was not observed. Conversely, mutants evolved to formic acid performed strongly when assayed with an inhibitor cocktail. This is particularly surprising given the presence of HMF is known to induce long lag times in *S. cerevisiae* [30]*.* These results favour the emergence of non-specific mutations to confer increased tolerance, such as the upregulation of genes such as *WHI2* encoding for a protein required to activate the general stress response and shown to increase tolerance to acetic acid and furfural both individually and in combination when overexpressed [31].

Lipid Production of Evolved Strains.

Though presence of inhibitors, whether as a single supplemented inhibitor or an inhibitor cocktail, is reported elsewhere to reduce biomass in oleaginous yeasts as shown here, their effect on lipid production varies [32–34]. Within *Rhodosporidium toruloides* for example [32]*,* no effect on lipid production was observed when cultured in the presence of six inhibitors, whereas the lipid accumulation of *C. curvatus* has been shown to reduce by 62% when in the presence of 1 g/L furfural [33]. In this study, the lipid production of the three groups of strains were affected by inhibitors in

different ways: the progenitor was unaffected, the formic acid evolved strains increased lipid production and the inhibitor cocktail evolved strains had their production reduced. Despite this, certainly the most striking result was the overall lipid production of inhibitor cocktail evolved strains increasing compared with the progenitor. Furthermore, as the formic acid evolved strains did not share this additional phenotype, it is likely that adaptation to the aldehyde inhibitors is responsible. An indication as to what adaptations may have occurred when evolving to a cocktail containing aldehyde inhibitors are given by how these compounds are processed by the cells. Within *S. cerevisiae,* furfural and HMF are both converted into less toxic compounds by NADPH-dependant enzymes [28,35] meaning evolution to media containing both may have selected for adaptations to increase cellular concentrations of NADPH for quicker inhibitor detoxification. Mutations or gene copy number amplifications of *POS5,* responsible for the conversion of NADH to NADPH within the mitochondria, or *ALD6* which converts $NADP^+$ into NADPH within the cytosol are therefore potential targets [36]. NADPH also plays a dual role within oleaginous organisms, functioning as a critical component in fatty acid synthesis. Despite it traditionally believed that NADPH used during fatty acid synthesis was derived solely from the activity of a cytosolic malic enzyme, recent studies have revealed that NADPH produced elsewhere within the cell can also be used to this end. These include the activities of glucose-6-phosphate dehydrogenase (G6PD) or 6-phosphogluconate dehydrogenase (PGD) in the pentose phosphate pathway, meaning that several alternative NADPH sources could be involved in the increased lipogenesis observed by these strains [37,38]. Given this, it is possible therefore that if a mutation has caused cellular levels of NADPH to be greater than that of the progenitor, particularly in conditions where inhibitors are not present, that fatty acid synthesis is providing a 'sink' for the excess NADPH. Furthermore, if these mutations were combined with a gene duplication to the lipid biosynthesis regulator *PDR16* as was observed within a *S. cerevisiae* strain evolved to fermentation inhibitors [29], then their synergistic effect may be responsible for this new phenotype.

An alternate explanation for increased lipid production within the multi inhibitor evolved strains could be due to an altered C:N response. Applying an experimental

design to assay C:N response of two evolved strains, 4x3 and F3, revealed distinct differences from its progenitor stage. In particular, high lipid accumulation by 4x3 was shown to occur over a broader range of C:N ratios assayed compared to the progenitor, with the reverse true for strain F3. As both strategies took place within nitrogen limited media, evolution to display a nitrogen limited, and therefore lipid rich, phenotype at lower ratios of carbon to nitrogen does not seem to have occurred due to the dissimilar C:N response by these two evolved strains. Again therefore it suggests that adaptation to the inhibitors themselves, rather than adaptation to nitrogen limited media, is responsible for the altered C:N response of 4x3. Indeed, it is reported that genes indicating nitrogen starvation were up regulated in *S. cerevisiae* in response to HMF and furfural supplemented media, with it further stated that NADPH flux usually directed towards ammonium assimilation under control conditions was instead redirected to detoxification of HMF and furfural [39].

Although overall lipid production was influenced by the presence of inhibitors (Figure 6), the FAME profile for all strains remained relatively unchanged (Table 1), agreeing with work by Hu et al. and Yu et al. [32,33]. What did affect the FAME profile however was culture age as shown within the non-sterile bioreactor cultures (Table 2). Under both inhibiting and non-inhibiting conditions, initially high levels of C16:0 within the first 48 hours of growth reduced and C18:1 and C16:1 levels increased as the cultures progressed. Here, it is likely that C16:0 fatty acids are either being converted enzymatically to C16:1 through the activity of desaturases such as *OLE1* [40], or further elongated (either before or after desaturation to C16:1) to C18:1. Maintaining a more saturated lipid, suitable as a palm oil substitute, would therefore require the cessation of culturing after 48 hours, or targeting enzymes such as *OLE1* for deletion. Whereas if a more monounsaturated product was required, for biodiesel production, then the yeast would need a longer time in the bioreactor.

The strain presented in this study appears unique amongst oleaginous organisms in producing a microbial oil high in C16:0, with *Cryptococcus curvatu* (28%), *R. glutinis* (18%) and *Y. lipolytica* (11%) having a notably lower proportion [41]. The narrow FAME profile of this organism, consisting almost exclusively of C16:0 and C18:1 fatty acids,

also appears contrary to other oleaginous yeast. The FAME profile of *Y. lipolytica* and *Trichosporon pullulans* for example is reported to consist 51% and 24% of C18:2 respectively, whilst *C. curvatus* and *R. graminis* produce 15% and 12% C18:0; FAME's *M. pulcherrima* appears to produce only in trace amounts [41,42].

This study emphasises the effectiveness of adaptive evolution to generate improved phenotypes in instances where genetic engineering is either not available, or, as with the aims of this study, difficult to approach with rational design due to the vast network of molecular mechanisms. By employing different strategies, though both successful with respect to increasing inhibitor tolerance, adaptive evolution can provide strain improvements beyond the initial aim.

4.7. References

[1] Sitepu, I.R.; Garay, L.A.; Sestric, R.; Levin, D.; Block, D.E.; German, J.B.; Boundy-Mills, K.L. Oleaginous yeasts for biodiesel: Current and future trends in biology and production. *Biotechnol. Adv.* 2014, *32*, 1336–1360.

[2] Whiffin, F.; Santomauro, F.; Chuck, C.J. Toward a microbial palm oil substitute: Oleaginous yeasts cultured on lignocellulose. *Biofuels, Bioprod. Biorefining* 2016.

[3] Long, Z.; Santomauro, F.; Fan, J.; Macquarrie, D.; Clark, J.; Chuck, C.J.; Budarin, V. Fast microwave-assisted acidolysis: a new biorefinery approach for the zero-waste utilisation of lignocellulosic biomass to produce high quality lignin and fermentable saccharides. *Faraday Discuss.* 2017.

[4] Fan, J.; Santamauro, F.; Budarin, V.; Whiffin, F.; Abeln, F.; Chantasuban, T.; Gore-Lloyd, D.; Henk, D.; Scott, R.J.; Clark, J.; et al. The additive free microwave hydrolysis of lignocellulosic biomass for fermentation to high value products. *J. Clean. Prod.* 2018, *198*, 776–784.

[5] Klinke, H.B.; Thomsen, A.B.; Ahring, B.K. Inhibition of ethanol-producing yeast and bacteria by degradation products produced during pre-treatment of biomass. *Appl. Microbiol. Biotechnol.* 2004, *66*, 10–26.

[6] Taylor, M.P.; Mulako, I.; Tuffin, M.; Cowan, D. Understanding physiological responses to pre-treatment inhibitors in ethanologenic fermentations. *Biotechnol. J.* 2012, *7*, 1169–1181.

[7] Radecka, D.; Mukherjee, V.; Mateo, R.Q.; Stojiljkovic, M.; Foulquié-Moreno, M.R.; Thevelein, J.M. Looking beyond Saccharomyces: The potential of non-conventional yeast species for desirable traits in bioethanol fermentation. *FEMS Yeast Res.* 2015, *15*.

[8] Lin, F.M.; Qiao, B.; Yuan, Y.J. Comparative proteomic analysis of tolerance and adaptation of ethanologenic Saccharomyces cerevisiae to furfural, a

lignocellulosic inhibitory compound. *Appl. Environ. Microbiol.* 2009, *75*, 3765–3776.
[9] Field, S.J.; Ryden, P.; Wilson, D.; James, S. a; Roberts, I.N.; Richardson, D.J.; Waldron, K.W.; Clarke, T. a Identification of furfural resistant strains of Saccharomyces cerevisiae and Saccharomyces paradoxus from a collection of environmental and industrial isolates. *Biotechnol. Biofuels* 2015, *8*, 33.
[10] Martín, C.; Marcet, M.; Almazán, O.; Jönsson, L.J. Adaptation of a recombinant xylose-utilizing Saccharomyces cerevisiae strain to a sugarcane bagasse hydrolysate with high content of fermentation inhibitors. *Bioresour. Technol.* 2007, *98*, 1767–1773.
[11] Dragosits, M.; Mattanovich, D. Adaptive laboratory evolution--principles and applications for biotechnology. *Microb. Cell Fact.* 2013, *12*, 64.
[12] Nannemann, D.P.; Birmingham, W.R.; Scism, R.A.; Bachmann, B.O. Assessing directed evolution methods for the generation of biosynthetic enzymes with potential in drug biosynthesis. *Future Med. Chem.* 2011.
[13] Stanley, D.; Fraser, S.; Chambers, P.J.; Rogers, P.; Stanley, G.A. Generation and characterisation of stable ethanol-tolerant mutants of Saccharomyces cerevisiae. *J. Ind. Microbiol. Biotechnol.* 2010, *37*, 139–149.
[14] Wright, J.; Bellissimi, E.; de Hulster, E.; Wagner, A.; Pronk, J.T.; Van Maris, A.J.A. Batch and continuous culture-based selection strategies for acetic acid tolerance in xylose-fermenting Saccharomyces cerevisiae. *FEMS Yeast Res.* 2011, *11*, 299–306.
[15] Van Maris, A.J.A.; Winkler, A.A.; Kuyper, M.; De Laat, W.T.A.M.; Van Dijken, J.P.; Pronk, J.T. Development of efficient xylose fermentation in saccharomyces cerevisiae: Xylose isomerase as a key component. *Adv. Biochem. Eng. Biotechnol.* 2007, *108*, 179–204.
[16] Wisselink, H.W.; Toirkens, M.J.; Berriel, M.D.R.F.; Winkler, A.A.; Van Dijken, J.P.; Pronk, J.T.; Van Maris, A.J.A. Engineering of Saccharomyces cerevisiae for efficient anaerobic alcoholic fermentation of L-arabinose. *Appl. Environ. Microbiol.* 2007, *73*, 4881–4891.
[17] Kurosawa, K.; Laser, J.; Sinskey, A.J. Tolerance and adaptive evolution of triacylglycerol-producing Rhodococcus opacus to lignocellulose-derived inhibitors. *Biotechnol. Biofuels* 2015, *8*, 76.
[18] Santamauro, F.; Whiffin, F.M.; Scott, R.J.; Chuck, C.J. Low-cost lipid production by an oleaginous yeast cultured in non-sterile conditions using model waste resources. *Biotechnol. Biofuels* 2014, *7*, 34.
[19] RStudio RStudio: Integrated Development for R 2015.
[20] Bligh, E.G.; Dyer, W.J. A rapid method of total lipid extraction and purification. *Can. J. Biochem. Physiol.* 1959, *37*, 911–917.
[21] Plesset, J.; Ludwig, J.R.; Cox, B.S.; McLaughlin, C.S. Effect of cell cycle position on thermotolerance in Saccharomyces cerevisiae. *J. Bacteriol.* 1987, *169*, 779–784.
[22] Smith, J.; van Rensburg, E.; Görgens, J.F. Simultaneously improving xylose fermentation and tolerance to lignocellulosic inhibitors through evolutionary engineering of recombinant Saccharomyces cerevisiae harbouring xylose isomerase. *BMC Biotechnol.* 2014, *14*, 41.
[23] Koppram, R.; Albers, E.; Olsson, L. Evolutionary engineering strategies to

enhance tolerance of xylose utilizing recombinant yeast to inhibitors derived from spruce biomass. *Biotechnol. Biofuels* 2012, *5*, 32.
[24] Dunham, M.J.; Badrane, H.; Ferea, T.; Adams, J.; Brown, P.O.; Rosenzweig, F.; Botstein, D. Characteristic genome rearrangements in experimental evolution of Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci.* 2002, *99*, 16144–16149.
[25] Brown, C.J.; Todd, K.M.; Rosenzweig, R.F. Multiple duplications of yeast hexose transport genes in response to selection in a glucose-limited environment. *Mol. Biol. Evol.* 1998, *15*, 931–942.
[26] Piper, P.; Mahé, Y.; Thompson, S.; Pandjaitan, R.; Holyoak, C.; Egner, R.; Mühlbauer, M.; Coote, P.; Kuchler, K. The Pdr12 ABC transporter is required for the development of weak organic acid resistance in yeast. *EMBO J.* 1998, *17*, 4257–4265.
[27] Tenreiro, S.; Rosa, P.C.; Viegas, C.A.; Sá-Correia, I. Expression of the AZR1 gene (ORF YGR224w), encoding a plasma membrane transporter of the major facilitator superfamily, is required for adaptation to acetic acid and resistance to azoles in Saccharomyces cerevisiae. *Yeast* 2000, *16*, 1469–1481.
[28] Petersson, A.; Almeida, J.R.M.; Modig, T.; Karhumaa, K.; Hahn-Hägerdal, B.; Gorwa-Grauslund, M.F.; Lidén, G. A 5-hydroxymethyl furfural reducing enzyme encoded by the Saccharomyces cerevisiae ADH6 gene conveys HMF tolerance. *Yeast* 2006, *23*, 455–464.
[29] Sato, T.K.; Liu, T.; Parreiras, L.S.; Williams, D.L.; Wohlbach, D.J.; Bice, B.D.; Ong, I.M.; Breuer, R.J.; Qin, L.; Busalacchi, D.; et al. Harnessing genetic diversity in saccharomyces cerevisiae for fermentation of xylose in hydrolysates of alkaline hydrogen peroxide-pretreated biomass. *Appl. Environ. Microbiol.* 2014, *80*, 540–554.
[30] Sehnem, N.T.; da Silva Machado, A.; Leite, F.C.B.; de Barros Pita, W.; de Morais, M.A.; Ayub, M.A.Z. 5-Hydroxymethylfurfural induces ADH7 and ARI1 expression in tolerant industrial Saccharomyces cerevisiae strain P6H9 during bioethanol production. *Bioresour. Technol.* 2013, *133*, 190–196.
[31] Chen, Y.; Sheng, J.; Jiang, T.; Stevens, J.; Feng, X.; Wei, N. Transcriptional profiling reveals molecular basis and novel genetic targets for improved resistance to multiple fermentation inhibitors in Saccharomyces cerevisiae. *Biotechnol. Biofuels* 2016, *9*, 1–18.
[32] Hu, C.; Zhao, X.; Zhao, J.; Wu, S.; Zhao, Z.K. Effects of biomass hydrolysis by-products on oleaginous yeast Rhodosporidium toruloides. *Bioresour. Technol.* 2009, *100*, 4843–4847.
[33] Yu, X.; Zheng, Y.; Dorgan, K.M.; Chen, S. Oil production by oleaginous yeasts using the hydrolysate from pretreatment of wheat straw with dilute sulfuric acid. *Bioresour. Technol.* 2011, *102*, 6134–6140.
[34] Chen, X.; Li, Z.; Zhang, X.; Hu, F.; Ryu, D.D.Y.; Bao, J. Screening of oleaginous yeast strains tolerant to lignocellulose degradation compounds. In Proceedings of the Applied Biochemistry and Biotechnology; 2009; Vol. 159, pp. 591–604.
[35] Heer, D.; Heine, D.; Sauer, U. Resistance of Saccharomyces cerevisiae to high concentrations of furfural is based on NADPH-dependent reduction by at least two oxireductases. *Appl. Environ. Microbiol.* 2009, *75*, 7631–7638.

[36] Miyagi, H.; Kawai, S.; Murata, K. Two sources of mitochondrial NADPH in the yeast Saccharomyces cerevisiae. *J. Biol. Chem.* 2009, *284*, 7553–7560.
[37] Zhang, Y.; Adams, I.P.; Ratledge, C. Malic enzyme: The controlling activity for lipid production? Overexpression of malic enzyme in Mucor circinelloides leads to a 2.5-fold increase in lipid accumulation. *Microbiology* 2007, *153*, 2013–2025.
[38] Chen, H.; Hao, G.; Wang, L.; Wang, H.; Gu, Z.; Liu, L.; Zhang, H.; Chen, W.; Chen, Y.Q. Identification of a critical determinant that enables efficient fatty acid synthesis in oleaginous fungi. *Sci. Rep.* 2015, *5*, 11247.
[39] Ask, M.; Bettiga, M.; Mapelli, V.; Olsson, L. The influence of HMF and furfural on redox-balance and energy-state of xylose-utilizing Saccharomyces cerevisiae. *Biotechnol. Biofuels* 2013, *6*.
[40] Ledesma-Amaro, R.; Nicaud, J.M. Yarrowia lipolytica as a biotechnological chassis to produce usual and unusual fatty acids. *Prog. Lipid Res.* 2016, *61*, 40–50.
[41] Ageitos, J.M.; Vallejo, J.A.; Veiga-Crespo, P.; Villa, T.G. Oily yeasts as oleaginous cell factories. *Appl. Microbiol. Biotechnol.* 2011, *90*, 1219–1227.
[42] Beopoulos, A.; Cescut, J.; Haddouche, R.; Uribelarrea, J.L.; Molina-Jouve, C.; Nicaud, J.M. Yarrowia lipolytica as a model for bio-oil production. *Prog. Lipid Res.* 2009, *48*, 375–387.

This declaration concerns the article entitled:									
Enhanced inhibitor tolerance and increased lipid productivity through adaptive laboratory evolution in the oleaginous yeast *Metshnikowia pulcherrima.*									
Publication status (tick one)									
draft manuscript		**Submitted**		**In review**	X	**Accepted**		**Published**	
Publication details (reference)									
Candidate's contribution to the paper (detailed, and also given as a percentage).	The candidate contributed to/ considerably contributed to/predominantly executed the… Formulation of ideas: Directed evolution idea was formulated between myself and Daniel Henk, including the two-strategy approach. 75% Design of methodology: Methodology was designed by myself. 100% Experimental work: The batch culturing was performed by Yuxin Sze under my supervision. Follow up characterisation studies were performed by myself. 80% Presentation of data in journal format: Applied Microbiology and Biotechnology								
Statement from Candidate	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.								
Signed						**Date**			

5. Chapter 5: Whole genome sequencing of 4x3.

5.1. Preamble

In the previous chapter, adaptive laboratory evolution was used to improve the inhibitor tolerance of *M. pulcherrima* via two strategies; an inhibitor cocktail as the selective pressure, or a single inhibitor, formic acid. From this study, both strategies gave strains with increased inhibitor tolerance, however the strains evolved within inhibitor cocktail conditions had an added phenotype of increased lipid accumulation. The ease of whole genome sequencing and development of analysis software means that ALE studies yielding interesting strains can be followed up to find causative genetic changes. These causative changes can then be used to inform targets for genetic engineering into wildtype or lab strains.

Copy number mutations within the genomes of evolved strains commonly occur during the short timescales of ALE studies, suggesting this mechanism as an effective route to adaptation[1]. Copy number variants can be in the form of aneuploidies, i.e. the loss or gain of a copy relative to a diploid background, duplications or multi-copy variants[1]. CNVs generally influence the phenotype of evolved strains through the correlation between copy number, gene expression and protein abundance[2]. Furthermore, genes with increased copy number can improve overall fitness by gaining novel function[1]. That said, studies have shown that increased expression levels lead to only a minimal phenotypic effect in instances where the affected proteins function within a protein complex, due to stoichiometric imbalances[3]. CNVs can incorporate varying sizes of the strains genome, for instance chromosomal CNVs have been implicated in improving several industrially relevant phenotypes, including high pH tolerance, xylose utilisation, ethanol tolerance and *p*-coumaric and ferulic acid tolerance[1]. Alternatively, smaller regions of aneuploidy which effect individual genes or loci are also commonly observed within industrial strains of *S. cerevisiae*, for instance increased copy number of *SUC* and *MAL* loci within strains grown on sucrose and maltose respectively[1].

In this short chapter, whole genome sequencing was performed on the lead candidate strain from chapter 3, 4x3, to scope for causative genetic changes. Analysis will be broken down into CNVs which are likely to affect and improve inhibitor tolerance, and factors which could have caused the unexpected phenotype of increased lipid production.

5.2. Materials and Methods.

5.2.1. Whole genome sequencing

DNA extraction, library prep and sequencing was performed by Novogene (Beijing, China) and included standard read quality assessment and clean-up to remove barcode and adapter sequences. Sequencing was performed using Illumina HiSeq with paired ends and an average insert of 150 bp with an average sequencing depth of 50X per gene (approximately 10 million reads per strain). NCYC2580 reads were assembled as a reference for downstream comparisons using dipSPAdes[4], WebAugustus[5] for gene prediction using the *C. albicans* model allowing a single transcript for each gene, and BLAST against reference protein databases from the Candida Genome Order Browser[6] and one derived from a single completely sequenced *M. pulcherrima* strain (MpFS, lab strain used within this group). Reads for 4x3 was aligned against the reference using Burrows-Wheeler Aligner[7]. Genome Analysis Toolkit (GATK) DepthOfCoverage was used to obtain coverage for individual genes. Copy number variants were determined by subtracting coverage within NCYC2580 from 4x3, and instances where coverage was $\pm$ 15X different was determined as a copy number variant within the evolved strain. Data is represented as coverage within scaffolds, rather than matched against chromosomal position and coverage within plots is represented on a gene-by-gene basis. For analysis of genes with CNVs, the *S. cerevisiae* homologue for each *M. pulcherrima* gene was used. In some instances, no *S. cerevisiae* homologue is available for the *M. pulcherrima* gene, meaning analysis of gene function could not be performed.

5.3. Results and Discussion.

5.3.1. Copy Number Variants within 4x3.

The strain 4x3 was generated through adaptive evolution in the presence of 5-HMF, furfural, acetic acid and formic acid from progenitor strain NCYC2580. During this study, 4x3 was phenotypically characterised and found to have increased growth rate and decreased lag time in the presence of these inhibitors. 4x3 was also found to accumulate almost double the lipid of the progenitor (41% versus 22%). Whole genome sequencing was performed on 4x3 and progenitor to identify genetic changes and prospect for causative mutations giving the described phenotypes. To compare overall genetic structures, depth of coverage analysis was performed on a gene by gene basis. NCYC2580 per gene coverage count was subtracted from 4x3 coverage to identify regions of the genome which had CNVs relative to the progenitor, where positive values indicate increased CN and negative decreased. Genes in 4x3 which had coverage $\pm$ 15X coverage (from a baseline of 50X coverage for a gene with two copies) relative to progenitor were determined to be a CNV. CNVs derive predominantly from errors in mechanisms of DNA repair and can implicate different lengths of the genome depending on the mechanism. Homologous recombination for example can generate smaller regions of duplication or deletions, spanning one or a few genes, whilst fork stalling and template switch can lead to regions of 20 kb being incorrectly amplified[2,8]. As increased gene CN number usually has a correlative effect on gene expression, the presence of amplified genes in adaptively evolved strains are likely to be as a result of selective pressure[2]. Plotted data shows a wide distribution of CNVs across the genome, with instances of both single gene amplifications/deletions or changes across much larger regions of the genome (Figure 1). On the whole, CNVs, whether amplifications or reductions, were on the order of a single copy relative to the diploid background (i.e. diploid coverage yielded approximately 50X coverage, meaning $\pm$ 25X coverage is the equivalent of 1 or 3 copies). There was also no great bias towards a greater number of genes with increased or decreased CN (Table 1).

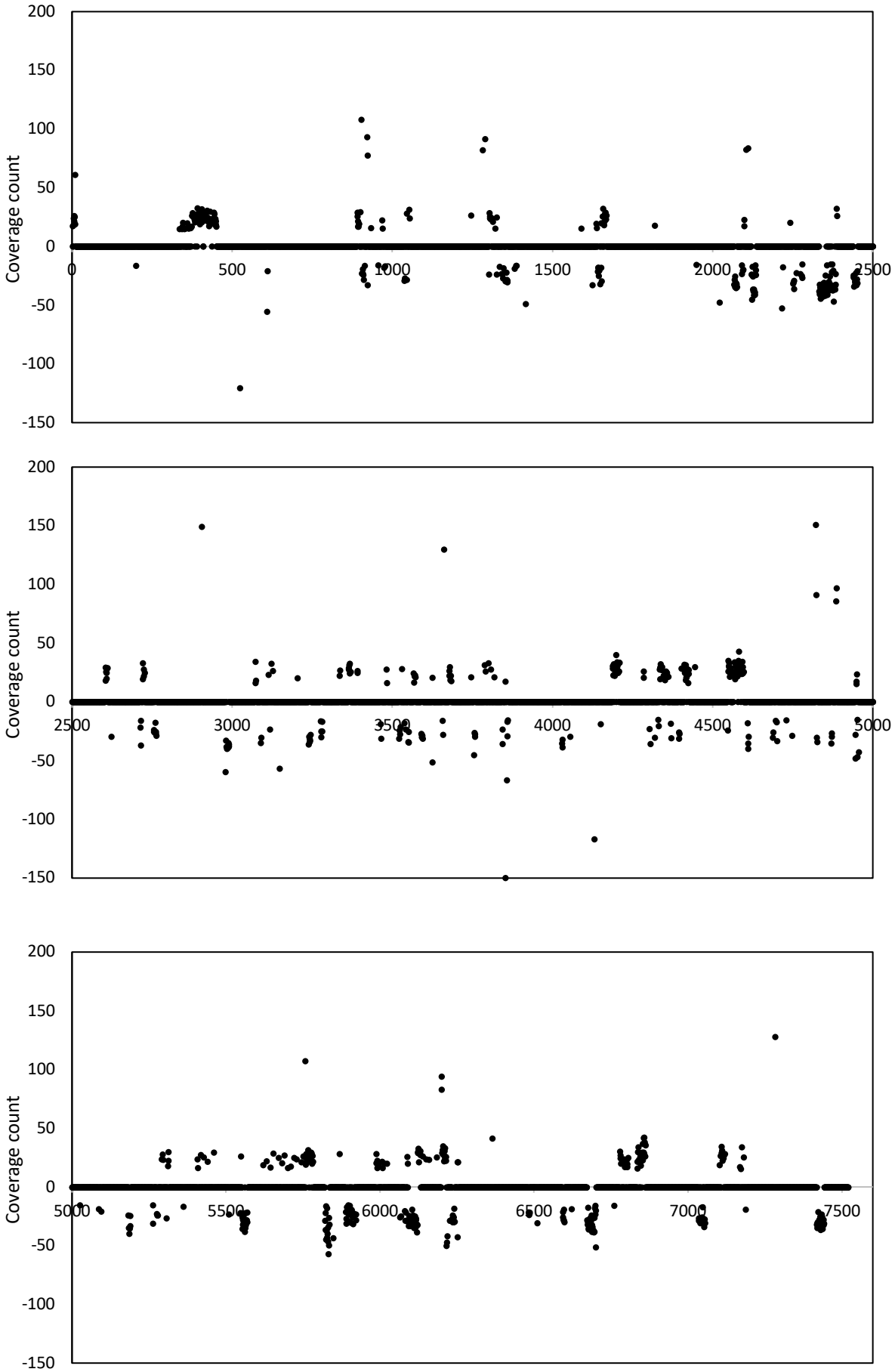

Coverage count
200
150
100
50
0
-50
-100
-150
0
500
1000
1500
2000
2500
Coverage count
200
150
100
50
0
-50
-100
-150
2500
3000
3500
4000
4500
5000
Coverage count
200
150
100
50
0
-50
-100
-150
5000
5500
6000
6500
7000
7500

Figure 1. Differently covered 4x3 genes with respect to NCYC2580 progenitor. Depth of coverage for each gene was determined for both strains. Gene coverage for NCYC2580 was subtracted from 4x3, identifying copy number variants whereby if coverage is positive the gene has higher copy number in 4x3 and lower copy number if negative. Genes which vary in coverage count by < 15X were determined to have no copy number difference, and are represented as zero. Data represents 7520 genes, split across three parts.

Table 1. Total number of 4x3 genes with CNVs relative to NCYC2580. Total number of genes which had > ± 15 coverage count relative to the progenitor. Total genes are further divided by the presence or absence of a S. cerevisiae homologue.

CVN Genes	Sc Gene Homologue	No Sc Homologue	Total
Increased	326	88	414
Decreased	353	92	445

Using the *S. cerevisiae* homologues, 4x3 genes with altered CN relative to the progenitor were pulled out to find those which may explain the observed phenotypic effects. Although there are over 850 genes within this criteria, instances where large sections of the genome are affected by a single event will mean that many genes will have altered CN due to their proximity to a gene with an evolutionary benefit, and may not have an effect themselves. There are two key phenotypes to investigate within 4x3; increased tolerance to fermentation inhibitors and increased lipid production.

5.3.2. Single copy variants

5.3.2.1. CNVs relating to inhibitor tolerance.

Present within the inhibitor cocktail used during the ALE study were two acid inhibitors, formic and acetic, and two aldehyde inhibitors, furfural and 5-HMF. As 4x3 did not have a significantly increased growth rate relative to the progenitor in an uninhibited, glucose-based media, it was previously hypothesised that mutations to increase glycolytic efficiencies were not likely in increasing fermentation inhibitor tolerance. Rather, it was anticipated that mutations would be specific to genes relating directly to the metabolism/stress response to these specific inhibitors

present. In total, four genes, including two plasma membrane transport proteins and two stress induced transcriptional activators, were identified with key functions relating to these inhibitors, as summarised in Table 2.

Table 2. Genes related to fermentation inhibitor tolerance with increased copy number in 4x3 relative to NCYC2580.

Gene	Function
PDR15	Plasma membrane ATP binding cassette (ABC) transporter. Strongly induced by various stress conditions including weak acid. Requires Msn2[9].
STB5	Transcriptional activator. Positively regulates NADPH producing enzymes including Ald6 which is implicated in aldehyde tolerance[10].
MSN2	Transcriptional activator. Overexpression confers furfural resistance and increases fermentation rate.[11] Regulates PDR15[9].
FLR1	Plasma membrane protein involved in multidrug resistance. Overexpression gave enhanced resistance to 5-HMF[12].

Although overexpression studies would be required to confirm the roles of these four genes in increasing inhibitor tolerance within 4x3, previous studies within *S. cerevisiae* give a strong likelihood of their role within the evolved phenotype of this strain. Expression of the ABC plasma membrane *PDR15* is strongly induced by weak acid, and increased CN of other transporters within this family, *PDR16* to *PDR18,* has conferred increased survival in high concentrations of ferulic acid[9,13]. Strongly increased induction of the *FLR1* plasma membrane transporter has been shown to occur in the presence of furfural, suggesting a role in their export from the cell, and overexpression of this gene has conferred enhanced resistance to 5-HMF[12,14]. Expression of *FLR1* was however unaffected by the presence of carboxylic acids, including formic and acetic acid[14]. Transcriptional activator *STB5* positively regulates the expression of NADPH producing enzymes, including Ald6[10]. Ald6 has been implicated in the direct oxidation of furfural and 5-HMF, generating a supply of NADPH within the reaction[15]. *STB5* also positively regulates *ADH6*, which when overexpressed in *S. cerevisiae* generated a strain within increased 5-HMF conversion[16]. However, although Adh6 was capable of using NADH, enzymatic activity was far higher when using NADPH as a co-factor[16]. Studies have identified

that overexpression of *MSN2* led to resistance to furfural, although unlike the other genes indicated, this transcriptional activator known to play a wider, major role in the yeast general stress response regulating hundreds of genes related to oxidative, temperature, pH, osmotic and nutrient stress[11,17,18]. Genome wide analysis studies have also associated Msn2 as a potential regulator factor in response to acetic acid[19].

5.3.2.2. CNVs relating to increased lipid production.

Genetic effects leading to increased lipid production was an unexpected outcome from the ALE study. It was hypothesised previously that increased lipid production was a consequence of adaptation to aldehyde inhibitors, as strains evolved to formic acid had reduced, rather than increased lipid accumulation. From this it was proposed that adaptations to increase cellular levels of NADPH required for detoxification of aldehyde inhibitors was also providing additional NADPH for lipid biosynthesis. This idea is supported as oleaginous organisms are known to have higher cellular concentrations of NADPH than non-oleaginous[20]. Finding increased CN of *STB5,* which positively regulates the expression of NADPH generating genes gives an indication that this hypothesis could be valid, especially as overexpressing *STB5* has been shown to increase cellular NADPH levels in *S. cerevisiae*[21].

As previously discussed, selection for increased *STB5* CN was likely as a result of positive regulation of Ald6 and Ahd6, which share activities in detoxification of furfural and 5-HMF. Though Ald6 is NADPH producing, studies have shown that the other Stb5 related target, Adh6, preferentially consumes NADPH as a co-factor when acting on aldehyde inhibitors[16]. This supports the observed decrease in lipid production by 4x3 when grown in the presence of inhibitors, when NADPH may be required by Adh6 activity. One caveat to that theory however is that co-factor preference between NADH and NADPH in reducing 5-HMF is strain dependent in *S. cerevisiae*, meaning usage in *M. pulcherrima* is difficult to determine at this time without dedicated study[16].

To find other potential causative genetic changes, genes with increased CN in 4x3 were examined for their roles in lipid accumulation. Here, *FAS1,* encoding an enzyme within the fatty acid synthetase (FAS) complex responsible for long chain fatty acids, was identified. Expression of *FAS1* has been shown to be upregulated 8-fold in conditions conducive to lipid accumulation[22]. Furthermore, a *S. cerevisiae* strain with multiple copies of *FAS1* resulted in a 10-fold expression increase of the second enzyme within the FAS complex; *FAS2*[23].

5.3.3. Regions of high variance.

The above considered the implications of genes which had increased in coverage by a single copy relative to the progenitor. In addition to these changes, ALE can also generate multi-copy variations, or change the copy number of genes which were previously represented in multiple copies. Figure 1 showed that the vast majority of CNVs were $\pm$ 25X coverage relative to the progenitor, however some genes had coverage greater than this difference, indicating such multicopy CVNs.

Two genes with a potential phenotypic effect with coverage suggesting higher than a single copy increase*. POR1*, now with approximately 100X coverage suggesting four copies, is a gene necessary for mitochondrial function has been identified within a genome-wide study to be implicated within in the acetic acid stress response[24]. Second, *ATF2*, an alcohol acetyltransferase implicated in tolerance to lignocellulose derived fermentation inhibitors, was identified with a similar level of coverage to *POR1*[25].

There were also instances of genes with four copies reverting back to two in the evolved strain. With respect to increased lipid production, reduced copy number of *AYR1* may play a contributing factor, with this gene identified as a triacylglycerol lipase within *S. cerevisiae*. Knockout strains for this gene had approximately 60% reduced conversion of triacylglycerol than the wildtype[26]. Interestingly, copy number of *PMA1*, encoding a membrane proton pump, was also reduced. Overexpression of this gene enhances tolerance to various types of stress within S. cerevisiae, including

acetic acid, making it an unusual to have reduced copy number in this evolved strain[27].

5.4. Conclusion.

The results from this initial study identify potential causative genetic changes within this evolved strain. That said, without comparative genomics of all the strains evolved within this study, or genetic studies to confirm a phenotypic effect, these results can only provide hypotheses. As improved lipid production was not a trait selected for during the ALE study, it was likely that this phenotype was a consequence of adaptation to the fermentation inhibitor cocktail. Being a key building block for fatty acid synthesis, it was hypothesised previously that increased cellular NADPH levels as a result of adaptation could be being re directed to fatty acid synthesis. In that regard, increased copy number was found of *STB5*, a transcriptional activator of NADPH producing enzymes, including *ALD6*, an enzyme responsible for detoxifying aldehyde inhibitors. The ABC plasma membrane, *PDR15,* was also found with increased copy number in 4x3 and is a good candidate for increased tolerance to the acid inhibitors.

5.5. References

1. Gorter de Vries, A. R., Pronk, J. T. & Daran, J. M. G. Industrial relevance of chromosomal copy number variation in Saccharomyces yeasts. *Appl. Environ. Microbiol.* (2017). doi:10.1128/AEM.03206-16
2. Steenwyk, J. L. & Rokas, A. Copy number variation in fungi and its implications for wine yeast genetic diversity and adaptation. *Frontiers in Microbiology* (2018). doi:10.3389/fmicb.2018.00288
3. Veitia, R. A., Bottani, S. & Birchler, J. A. Cellular reactions to gene dosage imbalance: genomic, transcriptomic and proteomic effects. *Trends in Genetics* (2008). doi:10.1016/j.tig.2008.05.005
4. Safonova, Y., Bankevich, A. & Pevzner, P. A. DIPSPADES: Assembler for highly polymorphic diploid genomes. in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (2014). doi:10.1007/978-3-319-05269-4_21
5. Stanke, M. & Morgenstern, B. AUGUSTUS: A web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* (2005). doi:10.1093/nar/gki458
6. Byrne, K. P. & Wolfe, K. H. The Yeast Gene Order Browser: Combining curated homology and syntenic context reveals gene fate in polyploid

species. *Genome Res.* (2005). doi:10.1101/gr.3672305
7. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* (2010). doi:10.1093/bioinformatics/btp698
8. Slack, A., Thornton, P. C., Magner, D. B., Rosenberg, S. M. & Hastings, P. J. On the mechanism of gene amplification induced under stress in Escherichia coli. *PLoS Genet.* (2006). doi:10.1371/journal.pgen.0020048
9. Wolfgert, H., Manmun, Y. M. & Kuchler, K. The Yeast Pdr15p ATP-binding Cassette (ABC) Protein Is a General Stress Response Factor Implicated in Cellular Detoxification. *J. Biol. Chem.* (2004). doi:10.1074/jbc.M311282200
10. Larochelle, M., Drouin, S., Robert, F. & Turcotte, B. Oxidative Stress-Activated Zinc Cluster Protein Stb5 Has Dual Activator/Repressor Functions Required for Pentose Phosphate Pathway Regulation and NADPH Production. *Mol. Cell. Biol.* (2006). doi:10.1128/MCB.02450-05
11. Sasano, Y. *et al.* Overexpression of the yeast transcription activator Msn2 confers furfural resistance and increases the initial fermentation rate in ethanol production. *J. Biosci. Bioeng.* (2012). doi:10.1016/j.jbiosc.2011.11.017
12. Alriksson, B., Horváth, I. S. & Jönsson, L. J. Overexpression of Saccharomyces cerevisiae transcription factor and multidrug resistance genes conveys enhanced resistance to lignocellulose-derived fermentation inhibitors. *Process Biochem.* (2010). doi:10.1016/j.procbio.2009.09.016
13. Sato, T. K. *et al.* Harnessing genetic diversity in saccharomyces cerevisiae for fermentation of xylose in hydrolysates of alkaline hydrogen peroxide-pretreated biomass. *Appl. Environ. Microbiol.* **80,** 540–554 (2014).
14. Westman, J. O., Taherzadeh, M. J. & Franzén, C. J. Inhibitor tolerance and flocculation of a yeast strain suitable for second generation bioethanol production. *Electron. J. Biotechnol.* (2012). doi:10.2225/vol15-issue3-fulltext-8
15. Park, S. E. *et al.* Expression of aldehyde dehydrogenase 6 reduces inhibitory effect of furan derivatives on cell growth and ethanol production in Saccharomyces cerevisiae. *Bioresour. Technol.* **102,** 6033–6038 (2011).
16. Petersson, A. *et al.* A 5-hydroxymethyl furfural reducing enzyme encoded by the Saccharomyces cerevisiae ADH6 gene conveys HMF tolerance. *Yeast* **23,** 455–464 (2006).
17. Causton, H. C. *et al.* Remodeling of Yeast Genome Expression in Response to Environmental Changes. *Mol. Biol. Cell* (2001). doi:10.1091/mbc.12.2.323
18. Sadeh, A., Movshovich, N., Volokh, M., Gheber, L. & Aharoni, A. Fine-tuning of the Msn2/4-mediated yeast stress responses as revealed by systematic deletion of Msn2/4 partners. *Mol. Biol. Cell* (2011). doi:10.1091/mbc.E10-12-1007
19. Mira, N. P., Becker, J. D. & Sá-Correia, I. Genomic Expression Program Involving the Haa1p-Regulon in *Saccharomyces cerevisiae* Response to Acetic Acid. *Omi. A J. Integr. Biol.* (2010). doi:10.1089/omi.2010.0048
20. Wasylenko, T. M., Ahn, W. S. & Stephanopoulos, G. The oxidative pentose phosphate pathway is the primary source of NADPH for lipid overproduction from glucose in Yarrowia lipolytica. *Metab. Eng.* (2015).

doi:10.1016/j.ymben.2015.02.007
21. Kim, J. E., Jang, I. S., Sung, B. H., Kim, S. C. & Lee, J. Y. Rerouting of NADPH synthetic pathways for increased protopanaxadiol production in Saccharomyces cerevisiae. *Sci. Rep.* (2018). doi:10.1038/s41598-018-34210-3
22. Schweizer, M. *et al.* The pentafunctional FAS1 gene of yeast: its nucleotide sequence and order of the catalytic domains. *MGG Mol. Gen. Genet.* (1986). doi:10.1007/BF00422073
23. Wenz, P., Schwank, S., Hoja, U. & Schüller, H. J. A downstream regulatory element located within the coding sequence mediates autoregulated expression of the yeast fatty acid synthase gene FAS2 by the FAS1 gene product. *Nucleic Acids Res.* (2001). doi:10.1093/nar/29.22.4625
24. Mira, N. P., Palma, M., Guerreiro, J. F. & Sá-Correia, I. Genome-wide identification of Saccharomyces cerevisiae genes required for tolerance to acetic acid. *Microb. Cell Fact.* (2010). doi:10.1186/1475-2859-9-79
25. Adeboye, P. T., Bettiga, M. & Olsson, L. ALD5, PAD1, ATF1 and ATF2 facilitate the catabolism of coniferyl aldehyde, ferulic acid and p-coumaric acid in Saccharomyces cerevisiae. *Sci. Rep.* (2017). doi:10.1038/srep42635
26. Ploier, B. *et al.* Screening for hydrolytic enzymes reveals Ayr1p as a novel triacylglycerol lipase in Saccharomyces cerevisiae. *J. Biol. Chem.* (2013). doi:10.1074/jbc.M113.509927
27. Lee, Y. *et al.* Overexpression of PMA1 enhances tolerance to various types of stress and constitutively activates the SAPK pathways in Saccharomyces cerevisiae. *Appl. Microbiol. Biotechnol.* (2017). doi:10.1007/s00253-016-7898-5

6. The oleaginous yeast *M. pulcherrima* improves xylose utilisation through different routes during adaptive laboratory evolution.

6.1. Commentary.

The previous two chapters followed the generation of fermentation inhibitor tolerant strains through adaptive evolution. From this, a candidate strain, 4x3, was selected for characterisation within bioreactors and genomic analysis to prospect for causative changes behind acquired phenotypes.

The following chapter takes this candidate strain and undertakes further adaptive evolution in attempt to improve metabolism of xylose. Although *M. pulcherrima* is able to naturally metabolise this carbon source, growth is far slower compared with other substrates such as glucose, maltose and cellobiose. Using a carbon source as the selective pressure will also provide a different adaptive backdrop to the previous ALE experiment. This will allow a wider comparison to be made between the routes of adaptation when there is a potentially lethal selective pressure (fermentation inhibitors) versus a non-lethal selective pressure in xylose.

The oleaginous yeast *M. pulcherrima* improves xylose utilisation through different routes during adaptive laboratory evolution.

Robert H. Hicks[a*], Jenny Thomas[b], Felix Abeln[c], Christopher J. Chuck[c], Roderick J Scott[b], David J. Leak[b], Daniel A. Henk[b].

[a*] Corresponding author. Centre for Doctoral Training in Sustainable Chemical Technologies, Department of Biology and Biochemistry, University of Bath, Bath, BA2 7AY, United Kingdom

[b] Department of Biology and Biochemistry, University of Bath, Bath, BA2 7AY, United Kingdom

[c] Department of Chemical Engineering, University of Bath, Bath, BA2 7AY, United Kingdom

Abstract

Background:

The oleaginous yeast *Metschnikowia pulcherrima* has previously been investigated as a sustainable route to producing a microbial oil for use in fuels, foods and bulk chemicals. Though naturally capable of metabolising xylose, consumption and biomass production from this carbon sourced lags behind glucose and presents a potential weakness within an industrial bioprocess based on lignocellulosic feedstocks. At present, suitable genetic tools for this organism are not available meaning adaptive laboratory evolution was used as a tool for strain improvement.

Results:

Starting with five parallel cell lines, all evolved strains were shown to have significantly improved biomass when grown on xylose as the sole carbon source after adaptive evolution. Though all cell lines were successful, the degree of improvement varied between the strains, with ALEX-5 outperforming the other strains to produce 17x greater biomass than the progenitor under the same conditions. Evolutionary trade-offs were seen across all strains through reduced lipid production and reduced biomass when grown on glucose, whilst biomass improved on ribose, another five-carbon sugar. Whole genome sequencing revealed a large and differing degree of copy number variation, between strains, suggesting multiple evolutionary routes to phenotypic improvement. ALEX-5 was the only strain to have increased copy number of a gene directly involved in xylose metabolism, *XKS1*, which potentially implicates this gene as causative in the marked improvement seen in this strain. Interestingly, frequency of copy number variation did not correlate with improved xylose metabolism, suggesting a situation where unrelated genes could change their copy number at no detriment to overall fitness.

Conclusions:

Adaptive evolution of five parallel cell lines of *M. pulcherrima* produced five final strains with improved performance on xylose. Despite this, the degree of improvement and severity of deleterious evolutionary trade-offs differed drastically between the evolved strains, suggesting different adaptive routes. As lipid production was significantly reduced in all strains, this study highlights the effectiveness of adaptive evolution in achieving its intended aim but being unable to mitigate negative consequences to other phenotypic aspects.

Key words:
Adaptive Laboratory evolution, Microbial Oil, Lignocellulose, Xylose, Genomics, Biodiesel

6.2. Background

Oleaginous yeasts have great potential to provide sustainable alternatives to fossil fuel derived fuels and chemicals. Often with a highly saturated FAME profile, microbial oils can also enter the market as an alternative to vegetable oils such as palm oil, which due to their inclusion in biodiesel, have had a deleterious impact on forest coverage in tropical regions[1]. Challenges do exist in the commercialisation of microbial oils, namely processing complications associated with the lignocellulosic feedstocks needed to make such bioprocesses economically viable. The oleaginous yeast *Metschnikowia pulcherrima* has previously been investigated as a potential platform organism for the production of biofuels and chemicals from these feedstocks, due to its natural tolerance to fermentation inhibitors and ability to metabolise oligosaccharides[2]. Though *M. pucherrima* can metabolise the xylose present in these feedstocks, slow growth rates and incomplete utilisation of this carbon source could hinder its commercial potential.

Typically constituting between 10% and 25% of the biomass within hardwoods, agricultural residues and energy crops, xylan represents a large proportion of the sugars available for product formation[3]. To ensure full conversion of this carbon source, research has focussed on either yeasts naturally able to metabolise xylose, or used rational engineering to give *Saccharomyces cerevisiae* novel pathways for xylose utilisation. Though the yeast used within this study, *Metschnikowia pulcherrima*, is able to grow on xylose as a sole carbon source, final biomass production lags behind that obtained from glucose[4]. Unlike *S. cerevisiae* and other non-conventional yeasts, genetic tools are not available for *M. pulcherrima* meaning that adaptive evolution is currently the main tool for strain improvement. Evolutionary techniques have been successfully applied within engineered strains of *S. cerevisiae* to improve utilisation of xylose[5,6] and arabinose[7] as well as a non-engineered *S. cerevisiae* strain to improve glycerol utilisation[8]. Selection for improved utilisation of single carbon substrates has the potential however to generate evolutionary trade-offs, as strains redirect and rearrange cellular processes to these specific conditions[9]. A strategy of dynamic evolution has been successful in

alleviating these potential trade-offs, as demonstrated by their success when alternating growth on glucose, xylose and arabinose in single sugar batch format to improving the overall fermentation of mixed sugar cultures[10].

The availability of inexpensive whole genome sequencing has transformed the impact of directed evolution experiments, allowing for the genetic causality of improved industrially relevant trains to be determined. Through this process for example, SNPs to the *GUT1* and *UBR1* genes have been identified to confer increased glycerol metabolism, and SNPs within hexose transporters *GUT2*, *HXT5* and *HXT7* have generated glucose-insensitive xylose transporters[8,11]. Whole genome resequencing has also unveiled the role of copy number variants through whole-chromosome or segmental aneuploidies within evolution experiments[12,13].

In this study, an *M. pulcherrima* strain previously evolved to have increased fermentation inhibitor tolerance was grown in media containing xylose as the sole carbon source, in a sequential batch evolution strategy[14]. Evolved strains were phenotypically compared with the progenitor for improvements and sequenced to prospect for causative genetic changes.

6.3. Methods

6.3.1. Chemicals

Unless otherwise stated, chemicals were sourced from Sigma Aldrich and used without further purification.

6.3.2. Strains, strain maintenance and media

The *M. pulcherrima* strain '4x3' used as the progenitor in this study was obtained from a previous evolutional experiment, having derived originally from NCYC2580 sourced from the National Collection of Yeast Cultures[14]. Strains were maintained on malt extract agar (MEA) plates, and re-streaked on a fortnightly basis. For the preparation of overnight cultures, a single colony was inoculated into 10 mL SMB pH

5 (3% tryptic soy broth, 2.5% malt extract), and incubated at 25 °C with 200 rpm agitation. Optical densities throughout were measured at 595_{nm}. Media used within ALE study was a nitrogen limited broth (NLB) consisting of: Carbon source (e.g. xylose, or as described) 40 g/L, $(NH_4)_2SO_2$ 2 g/L, KH_2PO_4 7 g/L, $MgSO_4$ $7H_2O$ 1.5 g/L, $NaHPO_2$ 2 g/L and yeast extract 1 g/L. This media was autoclaved without the carbon source or $MgSO_4$, which was added separately after autoclaving individual stock solutions.

6.3.3. Evolution experiment

Five overnight cultures were grown in SMB at 25 °C, 200 rpm and diluted to an OD ~ 1 with PBS the following morning. The diluted cells were inoculated into five separate 10 mL NLB + xylose culture tubes, and left to incubate at 25 °C, 200 rpm for 48 hours. After this time, 100 µL of each culture was transferred into a fresh 10 mL culture tube of the same media, taking a glycerol stock at each transfer and storing at -80 °C. The 48-hour batch transfer system continued until Batch 11, where transfers were then performed after 24 hours. The end of the evolution experiment was concluded after approximately 1000 hours, at which point around 100 cells from all five cell lines were plated onto YNB + xylose agar plates. After 72 hours of growth, the largest colony was selected and transferred onto a fresh MEA plate for phenotypic analysis. Glycerol stocks were prepared for each selected colony for the preparation of cells for whole genome sequencing.

6.3.4. 96 well plate phenotypic analysis

To phenotypically analyse the resulting mutant cell lines, a 96 well plate format was used. For this, overnight cultures for each evolved strain, as well as the progenitor, were prepared as stated previously and diluted to an OD ~ 1 the following morning. To prepare the 96 well plate, 140 µL of media was added to the appropriate well, and to this, 10µL of each diluted culture was added. Media blanks were plated as negative controls. Cultures were incubated within a shaking plate reader with temperature maintained at 25 °C. Optical density was read at 30 minute intervals, for a period of 48 hours unless otherwise stated within figure legends. Max growth rates were

calculated using RStudio[15], using Ln(OD) values. Media composition was as described within figure legends.

6.3.5. Growth performance and lipid extraction

To phenotypically assess biomass and lipid production, overnight cultures for each strain were prepared as previous stated, and 500 µL of the diluted overnight culture was used to inoculate 10 mL of NLB + carbon source, as described within figure legends. Cultures were incubated for seven days at 25 °C, 200 rpm upon which 1 mL of culture was saved and placed in an oven overnight for dry weight analysis and the remaining 9 mL of culture used for lipid extraction.

The lipid extraction methodology is based upon that proposed by Bligh and Dyer [16], and modified to the following:

Upon completion of culturing, 9 mL of cells were centrifuged at 11.3 thousand RCF, resuspended in 1 mL PBS, and centrifuged again, discarding the supernatant and freezing cell pellets instantly in liquid nitrogen. Following this, frozen cell pellets were freeze dried at -40°C for a minimum of six hours. For the cell disruption, a preweighed amount of cell pellet (ideally within the range of 10 – 100 mg) was mixed with 10 mL 6M HCl and stirred at 80 °C for one hour. To extract lipids, 10 mL of chloroform:methanol (1:1) was added, and the mixture was stirred overnight at room temperature. To quantify the lipid weight, the lower chloroform phase was removed by hand with a glass pipette, avoiding the emulsion layer which forms between the chloroform and aqueous phase. The chloroform was then fully evaporated via rotary evaporator at 50 °C, and the remaining lipids were weighed to determine the lipid weight % of the cells.

6.3.6. HPLC analysis

For HPLC analysis, supernatant samples were diluted with ddH_2O where appropriate and filtered (0.22 µm). Glucose and xylose content was determined using Agilent 1260 Infinity LC system using an 300 × 7.8 mm Rezex™ RHM-Monosaccharide H+

(Phenomenex) HPLC column held at 80 °C and refractive index detector at 40 °C. A 5-point calibration curve was performed to quantify both sugars.

6.3.7. Whole genome sequencing and analysis.

DNA extraction, library prep and sequencing was performed by Novogene (Beijing, China) and included standard read quality assessment and clean-up to remove barcode and adapter sequences. Sequencing was performed using Illumina HiSeq with paired ends and an average insert of 150 bp with an average sequencing depth of 50X per gene (approximately 10 million reads per strain). 4x3 reads were assembled as a reference for downstream comparisons using dipSPAdes[17], WebAugustus[18] for gene prediction using the *C. albicans* model with and allowing a single transcript for each gene, and BLAST against reference protein databases from the Candida Genome Order Browser[19] and one derived from a single completely sequenced *M. pulcherrima* strain (MpFS, lab strain used within this group). Reads for evolved strains were aligned against 4x3 reference using Burrows-Wheeler Aligner[20]. Genome Analysis Toolkit (GATK) DepthOfCoverage was used to obtain coverage for individual genes within strains to determine copy number variants. Data is represented as coverage within contigs, rather than matched against chromosomal position. GATK was also used to obtain SNP variants within genes of evolved strains versus the progenitor gene sequence, and the SnpEff[21] tool was used to identify high impact variants from the identified SNPs for further analysis. High impact SNPs include several putative impacts including the loss of a start or stop codon, splice site mutation or frameshift mutations. For analysis of genes with CNV or high impact SNP, the *S. cerevisiae* homologue for each *M. pulcherrima* gene was used. In some instances, no *S. cerevisiae* homologue is available for the *M. pulcherrima* gene, meaning analysis of gene function could not be performed.

6.4. Results and discussion

6.4.1. Adaptive evolution.

Five clonal lineages were started in parallel by inoculation into 10 mL NLB + xylose and grown for 48 hours. All cultures were transferred after 48 hours growth by taking 100 μL of the previous culture through to a fresh 10 ml NLB + xylose culture tube. A steady increase in both the 24 hour and 48 hour OD occurred over the first 11 batches along with a reduction in lineage variability as shown by decreasing error bars (Figure 1). Due to the consistent increase in 48 hour transfer OD up to batch 11, transfers were then made after 24 hours to maintain a low cell population and encourage the selection of cell populations with increased growth rate and short lag time. This strategy was immediately followed by a reduced 24 hour OD in subsequent batches, most likely due to a reduced number of cells being inoculated at transfer. From batch 17 until batch 30 the cell lines showed evidence of adaptation as successive transfer ODs were, with a couple exceptions, higher than the previous. The variance of transfer OD across the five cell lines by the end of the experiment was minimal.

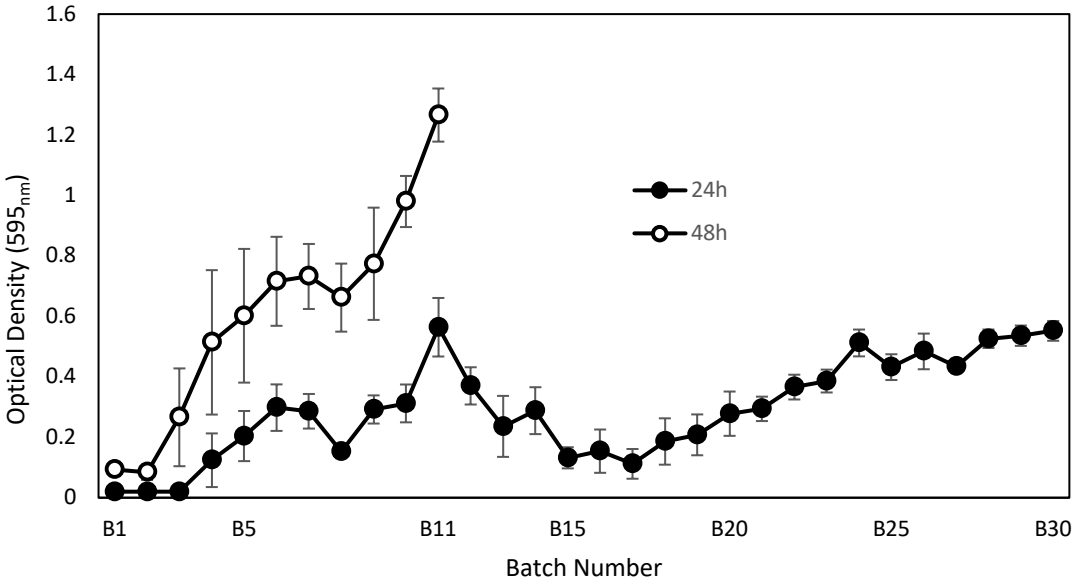


Figure 1. Overview of xylose adaptive evolution. Five clonal lineages were used to inoculate five NLB + xylose cultures. 24 h (filled circles) and 48 h (clear circles) optical density was tracked, and data presented represents the average and standard deviation across the five cultures. Batchers were transferred every 48 h until batch 11, where cells were then transferred every 24 h.

6.4.2. Phenotypic analysis of evolved strains

Adaptive evolution was concluded after batch 30, at which point approximately 100 cells were plated onto YNB + xylose agar plates. After 72 hours of growth, the largest colony was transferred to a fresh MEA plate and maintained for phenotypic analysis. The resulting evolved strain from each five cell lines was named ALEX-1, ALEX-2, ALEX-3, ALEX-4 and ALEX-5 respectively (Adaptive Laboratory Evolution Xylose 1, for example). Each evolved strain and the progenitor, 4x3, were first screened for growth on xylose as the sole carbon source in a 96 well plate (Figure 2). Here, although all the evolved strains reached a significantly ($p < 0.05$) higher final 48 hour OD than the progenitor, there were also large differences between the evolved strains themselves. ALEX-5 performed the best, reaching an average final OD of 0.47 compared to ALEX-4 which reached an average of 0.35. The variation between the evolved strains is perhaps unexpected due to the similar batch performance within the late stages of the evolution process (Figure 1).

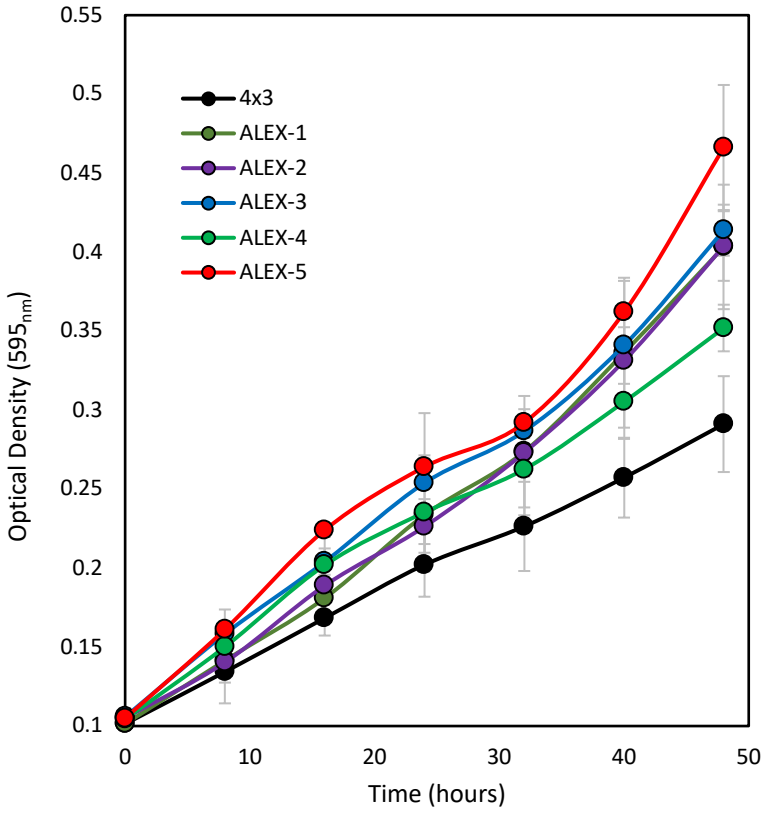


Figure 2. Growth curves of the progenitor and evolved strains in NLB + xylose. Each strain was assayed in triplicate within a 96-well plate. Data points represent

the average of triplicate wells, with error bars representing the standard deviation at each point.

As trade-offs are known to occur within single carbon source evolution experiments, growth in NLB + glucose was compared for all strains in the same format as for xylose. A significant difference between the progenitor and the evolved strains was again observed, this time with respect to the maximum growth rates achieved within this media (Figure 3a, 3b). All evolved strains exhibited slower growth when glucose was the sole carbon source, potentially due to the redirection of cellular metabolism and resources away from glucose during the evolution process, as observed elsewhere[22]. There was not as great a difference between the final ODs of the evolved strains and progenitor in this assay compared to NLB + xylose.

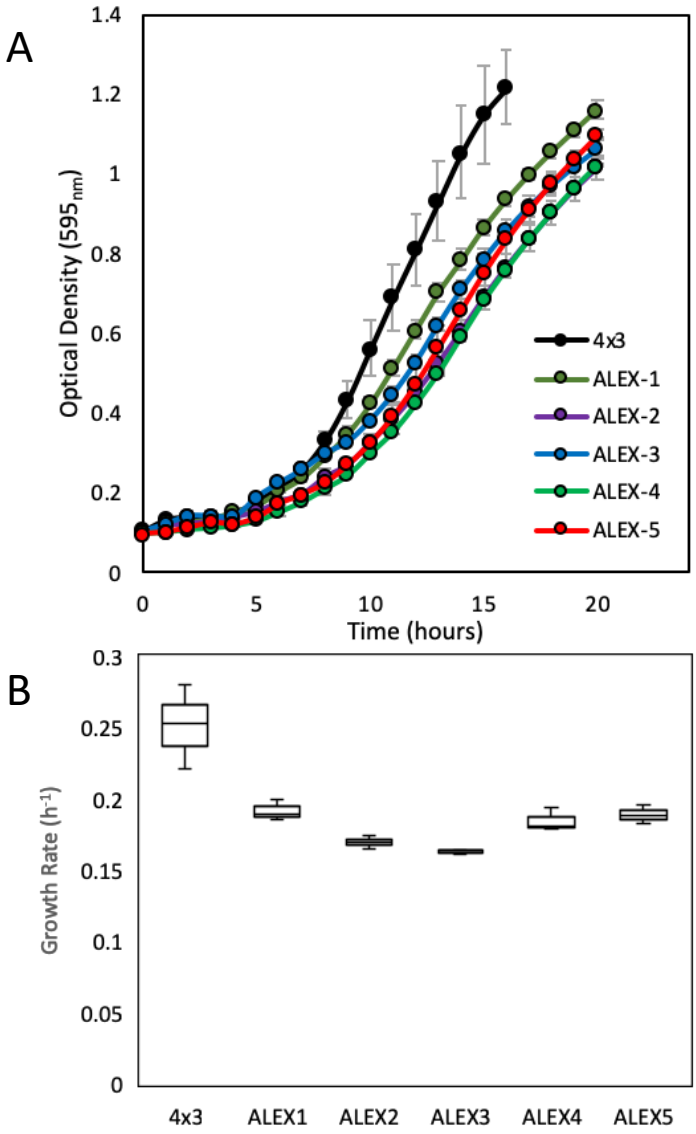


Figure 3. Growth curves and growth rates of the progenitor and evolved strains in NLB + glucose. Each strain was assayed in triplicate within a 96-well plate. Data points represent the average of triplicate wells, with error bars representing the standard deviation at each point. A: Growth curves and B) plotted growth rates. Max growth rates were calculated using R Studio[15].

The strain 4x3 was chosen for this study due to its high lipid accumulation (> 40%) but poor performance when growing on xylose as the sole carbon source. Earlier data presented showed how the evolved strains had improved performance on xylose within a 96-well plate format (Figure 2), at the expense of a decreased growth rate on glucose. To check for further evolutionary trade-offs, the lipid and biomass production of all strains was evaluated after seven days of growth on NLB + xylose and NLB + glucose (Figure 4A). In agreement with Figure 2, all ALEX evolved strains have increased biomass production on NLB + xylose (Figure 4A), however as indicated within the 96 well plate assay, variation was seen across the strains. ALEX-5 produced a remarkably high average biomass of 8.2 g/L; over double the second highest strain ALEX-4 at 3.3 g/L. ALEX-2 had the lowest biomass production of the evolved strains and was the only strain without statistically significant increased biomass ($p = 0.057$). In total, the biomass production of ALEX-5 has increased by over 17 times compared to the progenitor. A similar situation of high variance was presented when strains were grown in NLB + glucose (Figure 4b). As anticipated from previous results, the biomass from all evolved strains was below that of the progenitor. ALEX-2, the poorest performer on xylose, again yielded the lowest biomass, whilst ALEX-1 produced significantly more than the other evolved strains. Critically however, the lipid production of the evolved strains, across both carbon sources, decreased compared with the progenitor. This reduction is particularly evident in ALEX-5 and ALEX-3, with both strains now accumulating <10% lipid. Further results of interest include ALEX-1, which maintains >30% lipid accumulation when grown on glucose and together with the high biomass production, it appears this strain has the fewest phenotypic evolutionary trade-offs for glucose metabolism. ALEX-2, which as

previously mentioned produced the least biomass, does however accumulate 20% lipid in both conditions.

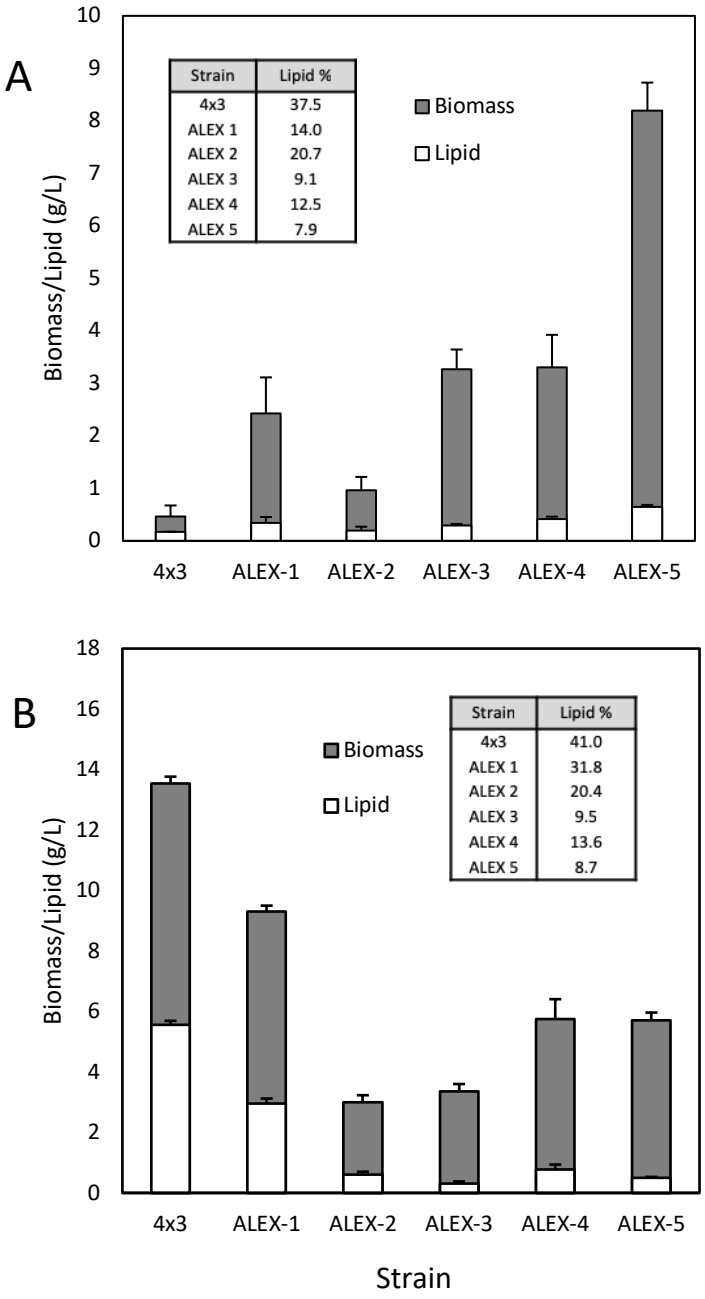


Figure 4. Lipid quantification and biomass of progenitor and evolved strains in NLB + xylose (A) and NLB + glucose (B). Triplicate samples were cultured for each strain for 7 days at 25 °C. Error bars represent the standard deviation of the mean. No error bars are present for 4x3 NLB +xylose lipid production as biomass production was too low to complete accurate lipid extractions, meaning all three samples were pooled for a single extraction.

Starting with a concentration of 40 g/L, xylose consumption by each strain was measured via HPLC (Figure 5). Results obtained closely mimic those presented in Figure 4, clearly highlighting ALEX-5's ability to utilise xylose in quantities far greater than the progenitor, and indeed all other evolved strains. Given that the media used is nitrogen limited, it is possible that ALEX-5 may be able to utilise a greater percentage of the available xylose if nitrogen hadn't become limiting. It is interesting in that case however that consumed xylose was not being routed into lipid accumulation, as could be expected when nitrogen limited and as occurs within 4x3. ALEX-2 and ALEX-3 both consumed < 10% of the xylose present, yet ALEX-3 produced significantly more biomass in this condition. This perhaps suggests that both strains are directing xylose uptake to different cellular processes, a hypothesis supported by ALEX-2 accumulating over double the quantity of lipid compared to ALEX-3. An important result throughout this phenotypic analysis is that despite the progenitor having very low biomass and xylose consumption, lipid quantification from this non-preferred substrate remains on par with that produced on glucose.

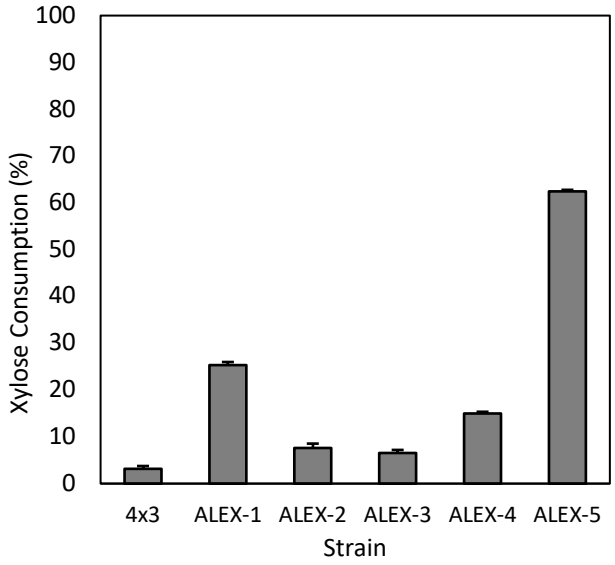


Figure 5. Percentage xylose consumption of progenitor and evolved strains.
Triplicate samples were cultured for each strain in NLB + xylose for 7 days at 25 °C. Xylose starting concentration was 40 g/L, remaining xylose was quantified by HPLC and represented as a percentage. Error bars represent the standard deviation of the mean.

Important phenotypic differences between the progenitor and evolved strains have been uncovered through single-sugar screening assays. Next, mixed sugar cultures was performed (Figure 6). Glucose-xylose mixed cultures generated increased biomass for ALEX-1-4 than within the single sugar screens, though reduced the final biomass of 4x3 compared to when grown in just glucose (Figure 6A). ALEX-5 was the only strain to reach a higher biomass when grown on xylose as the sole carbon source. ALEX-1 accumulated almost 40% lipid, bringing its performance in line with the progenitor, whilst the lipid production of the remaining strains was consistent with previous conditions. In contrast to data in Figure 5, xylose consumption when cells were grown in a mixed xylose/glucose culture was higher for all strains (Figure 6B). In fact, 4x3, ALEX-2 and ALEX-4, which all had <20% xylose consumption when present as the sole carbon source, had almost complete consumption when in the presence of glucose. Of the strains tested, only ALEX-3 failed to consume >90% of the available xylose. Although data is not available to confirm, it is likely these strains are preferentially consuming glucose, leading to an increased cell population, before xylose is then metabolised. A high starting cell population before commencing xylose metabolism would explain increased consumption compared with previous data, where initial cell numbers were far lower. Greater consumption of xylose when grown in the presence of glucose has been observed in *M. pulcherrima* elsewhere, possibly for the same reason[4].

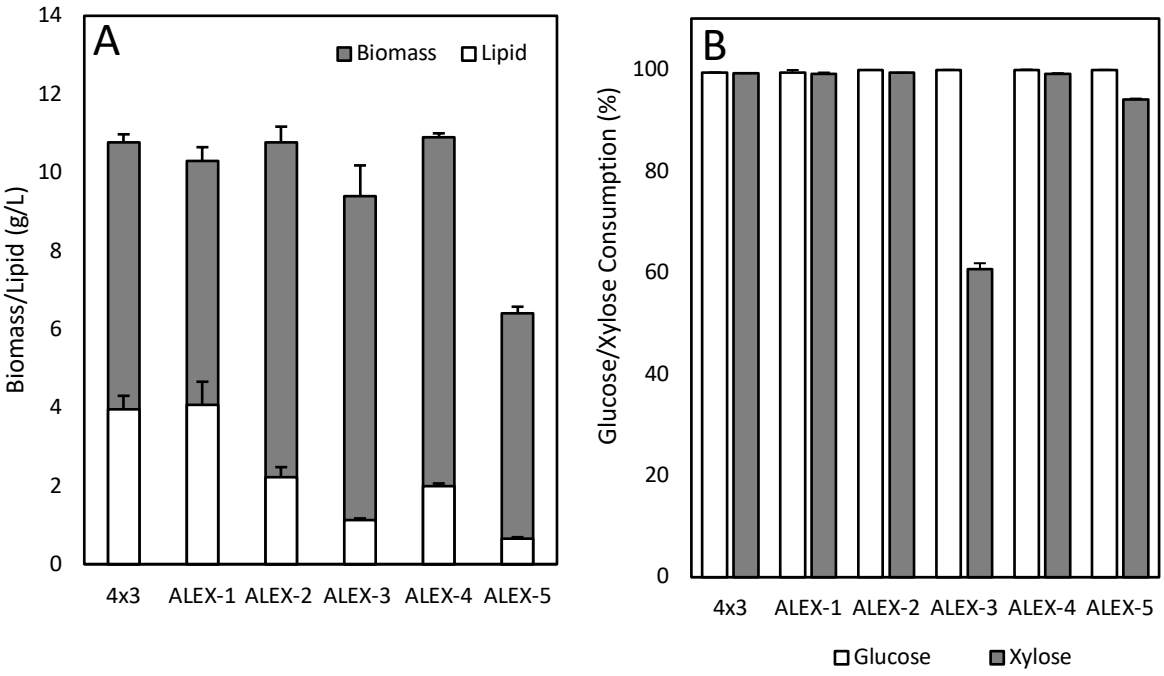


Figure 6. Lipid quantification, biomass and consumption of progenitor and evolved strains in NLB + xylose/glucose mix. Triplicate samples were cultured for each strain for 7 days at 25 °C. Total initial carbon was 40 g/L as previously used, 50/50 between glucose and xylose. Error bars represent the standard deviation of the mean. A – Lipid and biomass. B – percentage consumption of glucose/xylose after 7 days.

ALE derived metabolic trade-offs which reduce the capacity to grow on and accumulate lipid from glucose have been shown across all evolved strains in this study. In the same way that genetic rearrangement can have a deleterious effect on unrelated carbon metabolism, the reverse is possible when strains evolved to one carbon source are grown on a related carbon source due to the enrichment of multi-functional genes[23]. To test for this effect, growth on the pentose sugars ribose and arabinose was performed. With ribose, all strains except ALEX-1 had significantly higher biomass production than the progenitor, supporting a hypothesis of cross-adaptation to related carbon sources (Figure 7A). Though significant, the biomass improvement on ribose for strains ALEX-3, ALEX-4, and particularly, ALEX-5, were not as marked as they were when screened on xylose. The performance of ALEX-1 was particularly surprising, as the biomass and percentage xylose consumption shown in earlier figures suggested there should also have been improved biomass on this related sugar. In contrast to the ribose data, little improvement was seen in growth

on arabinose for all evolved strains (Figure 7B). In fact, as shown by the presentation of the data in optical density rather than g/L, biomass productivity from arabinose was very poor across all strains, correlating with previous studies on the same substrate for this yeast[4,24].

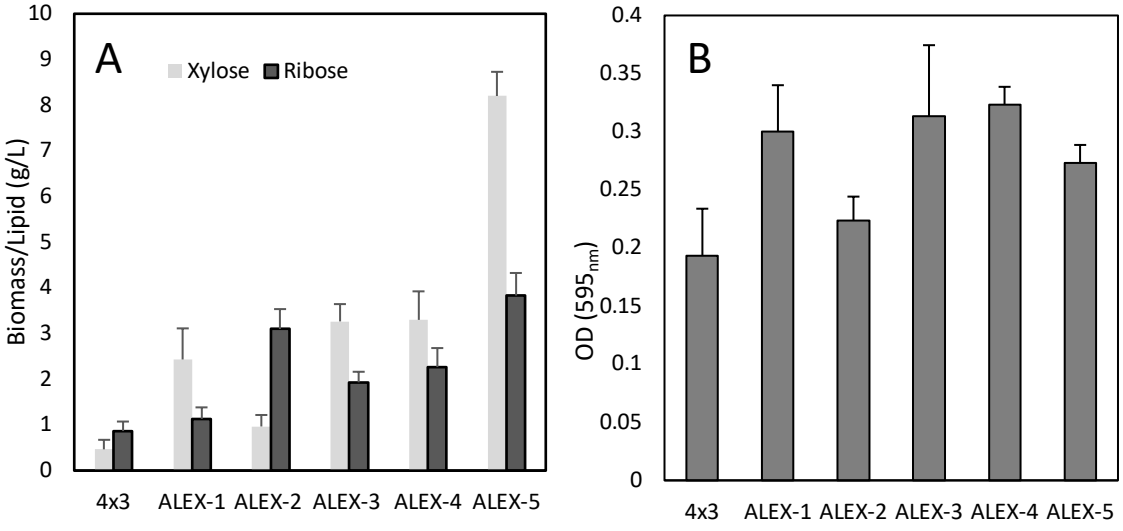


Figure 7. Biomass of progenitor and evolved strains in NLB + ribose and ODs in NLB + arabinose. Triplicate samples were cultured for each strain for 7 days at 25 °C. A – NLB + Ribose, B – NLB + Arabinose. Results from earlier NLB + xylose are overlaid in light grey for the NLB + ribose results. Error bars represent the standard deviation of the mean.

The loss and gain of carbon metabolism efficiencies through experimental evolution has proven to be evident, even after the short time span conducted within this study. The stand-out results are certainly the performance of ALEX-5, which now produces biomass over 17x that of the progenitor when grown on xylose, and the reduction in glucose biomass by all evolved strains and the effect this has on lipid production. A noteworthy caveat to the evolutionary trade off observed on glucose was the increased performance of the evolved strains on pentose sugar ribose and how this wasn't observed with arabinose, another five-carbon sugar. Cross-adaptation or cross-protection to a different stressor than the one selected for is not an uncommon outcome from directed evolution experiments in yeast and bacteria[25,26]. It is hypothesised, for example, that in adaptive evolution experiments to improve tolerance to osmotic or oxidative stress that beneficial evolutionary adaptations to the Environmental Stress Response (a response which induces the expression of ~300 stress defence genes) or general stress transcription factors like *MSN2* and *MSN4* are

responsible for providing cross adaption to stresses not selected for during the evolutionary regieme[27]. Though a different type of stress to oxidative, osmotic, temperature etc, it is possible that the cross-adaptation for improved ribose metabolism has occurred because of the shared metabolic pathways of these sugars within the pentose phosphate pathway (PPP)[28]. It is interesting therefore that only negligible improvements to arabinose metabolism was achieved considering it too feeds into the PPP. This suggests that any improvements or modifications within the PPP has not included the initial step of arabinose conversion, despite this step being often being catabolised within yeast by an aldose reductase also used within xylose metabolism[29].

6.4.3. Copy Number Variation within ALEX strains.

With a host of key phenotypes now identified, whole-genome resequencing was performed on all strains. Depth of coverage analysis was first performed to identify sequence-level changes, with the intention of attaching these to the observed phenotypic traits. Plotting individual gene coverage across the whole genome for each strain against coverage within 4x3 reveals both large variations between evolved strains and the progenitor, as well as variation between evolved strains (Figure 8, summarised in Table 1). Initial analysis considered contigs which had baseline diploid coverage (~50X coverage, i.e. two copies) within 4x3 but were now represented with altered coverage count. On a numerical basis, ALEX-2 has the highest number of copy number variants (CNVs), with 38 contigs now gaining an extra copy (~75X coverage, i.e. three copies). ALEX-1, the strain whose phenotype appeared least affected by the directed evolution experiment, had a comparatively low number of CNVs. ALEX-3 and ALEX-4 were the only strains to have contigs previously with two copies, now at one, three or four copies. There were a total of 17 contigs found in more than one strain, with seven of them found in all strains bar ALEX-1. ALEX-5, the strain with the most interesting xylose-phenotype, did not have a notably high amount of CNVs, but was however unique in that all CNVs were complete duplications (~100X coverage), meaning a total of 20 contigs were now

found with four copies. A complete list of contig CNVs and what strains they occur in are shown in Table S1.

With the exception of ALEX-5 and ALEX-1, though the latter contained comparatively few CNVs, the introduction of aneuploidy was the most common structural variation found within the evolved strains. Although aneuploidy typically reduces fitness relative to the progenitor by causing a downregulation in genes relating to cell growth and proliferation, it provides a population with genetic instability and therefore the opportunity for genetic variation in the face of a selective pressure[13,30]. Indeed, although aneuploidy is found within some lab strains of *S. cerevisiae*, its appearance is far more common within industrial or adaptively evolved strains where a specific and consistent environment has been present[30]. Gene expression levels within aneuploid sections of the genome, or indeed any CNV, generally correlate with copy number, and can have a significant effect on the observed phenotype. Though within the confines of a selective environment this gene dosage effect is beneficial, the generation of 'crippled' strains can be realised upon taking strains out of the environment, where the negative effects of aneuploidy have been otherwise hidden[31]. This effect is almost certainly at play within highly aneuploid strains ALEX-2,-3 and -4, and is even more likely given the clustering of metabolic genes within yeast chromosomes[32].

Table 1. Summary of gene coverage within evolved strains. Values refer to contigs which had 2x coverage (2 copies) within the progenitor but are now represented as described. Total amount of modified contigs and number of unique-to-strain copy number modifications are also tallied.

Coverage	4X	3X	1X	Total	Unique
ALEX-1	4	1	0	5	1
ALEX-2	0	38	0	38	24
ALEX-3	1	11	3	15	1
ALEX-4	1	16	2	19	4
ALEX-5	20	0	0	20	10

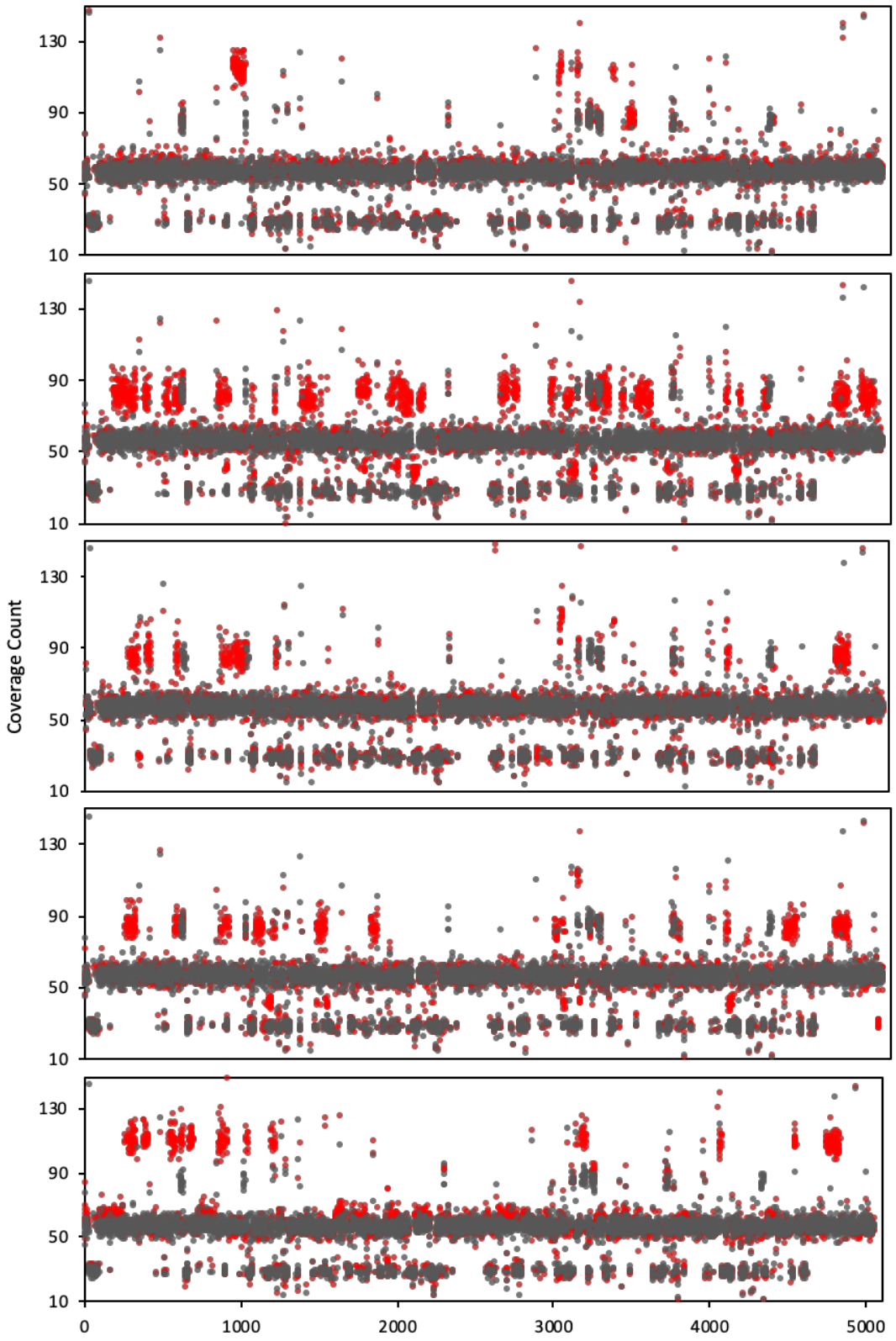


Figure 1. Gene-by-gene coverage count for ALEX strains against the progenitor. The gene-by-gene coverage count (genes plotted along the x-axis) for each evolved strain is plotted (red), with the progenitor (grey), 4x3, overlaid as reference. From top to bottom, ALEX-1, ALEX-2, ALEX-3, ALEX-4, ALEX-5.

Aside from regions of the evolved genomes which became aneuploid from a diploid state, there were also contigs within the 4x3 genome which started aneuploid with three copies and ended up either losing a copy to revert to diploid, or gain a further copy. In total, five contig regions with three copies in 4x3 had changed within the ALEX strains, in almost all cases a copy being lost to return coverage to baseline for a diploid (Table 2). Contig 83 stood out in particular as the only region which had reduced copy number in all evolved strains. Further analysis of this contig foun d several genes related to carbon metabolism; *GAL2*, a high affinity glucose/galactose transporter[33], *GAL3*, a transcriptional regulator induced by galactose[34] and *GAL7/GAL10* which are both enzymes within the galactose metabolic pathway[35,36] all reduced to diploid in all strains bar ALEX-3 which has just a single copy. Although not directly related to glucose metabolism, given this contig was aneuploid in the progenitor, and all evolved strains now have reduced biomass when grown on glucose, it is possible that the extra copy now lost in the evolved strains could have developed functional divergence[37]. This is particularly likely in the case of an extra *GAL2* copy, which could have gained specialist function for glucose transport. Contig 131 also had reduced copy number in all strains but ALEX-4, and within this contig was an NAD^+-dependent isocitrate dehydrogenase (*IDH1*), whose activity has been linked with lipid accumulation in oleaginous yeasts including *Rhodosporidium toruloides*[38]. Losing a copy of this gene could lead to reduced lipid accumulation within the ALEX strains, however as three copies of this contig remain in ALEX-4, it is unlikely to be the sole cause.

6.4.4. Further analysis of ALEX-5

In correlation with its unique phenotypic performance, ALEX-5 was unique in that all CNVs were duplications. Furthermore, out of these 20 CNVs, 10 were unique to this stain. As the expression of genes is reported to correlate with copy number, the CNVs unique to ALEX-5 were investigated[39]. Contig 116 immediately stands out due to it containing *XKS1*; a xyluokinase catalysing the final step of xylose metabolism before entering the PPP[40]. Overexpression of this gene has significantly increased xylose

utilisation in a previous study, correlating with the performance of ALEX-5[41]. As this was the only xylose metabolism specific gene within the duplicated contigs, it is difficult to pick out other genes and attribute them as the drivers for selection of duplicated regions with any certainty. For example, even though *GND2,* an enzyme catalysing an NADPH regenerating reaction within the pentose phosphate pathway[42], was within the duplicated contig 128, 15 other genes were also present on this contig.

Within duplicated regions, there were also genes which if overexpressed may have contributed to the reduced lipid production in this strain; *SNF2* and *POX1.* Disruption of *SNF2* within *S. cerevisiae* led to significantly increased lipid production, with it hypothesised that disruption of this gene altered carbon flux into storage lipid[43]. *POX1,* a fatty acyl-CoA oxidase involved in the β-oxidation of fatty acids, has been deleted in strategies to increase lipid production by preventing fatty acid degradation[44]. Given the low lipid accumulation of this strain, it is possible that ALEX-5 could be using increased rates of fatty acid β-oxidation to channel acetyl-CoA into the citric acid cycle for growth. These genes could therefore be investigated within knockout studies.

ALEX-5 was also the only strain in which genes were lost entirely, opposed to just having reduced copy number. The four genes lost were *GAL2, GAL3, GAL10 and YPS1*, and as presented earlier, the contig these genes sit within had reduced in copy number in all ALEX strains, suggesting further benefits to their complete removal. Though the *GAL* genes are specific for galactose metabolism, it was hypothesised earlier that the extra copy present within the progenitor could have gained functional divergence. Whether that is the case or not, it is clear that the loss of a gene copy in ALEX-1, -2, -3 and -4, and the complete loss of this gene cluster in ALEX-5, was selectively favourable within the evolutionary parameters.

Table 2. Summary of contigs with 3 copies in 4x3 which have now changed.
Values refer to contigs which had 3X copies in the 4x3 genome, but had changed to the coverage stated for each strain, where applicable.

Change from 3X copies in 4x3					
Contig	ALEX-1	ALEX-2	ALEX-3	ALEX-4	ALEX-5
131	2X	2X	2X	-	2X
113	-	-	2X	-	4X
47	-	-	2X	-	4X
83	2X	2X	1X	2X	2X*
46	-	-	-	4X	-
*Although most genes returned to diploid, four genes were lost from this contig.					

6.4.5. Frequencies of CNV and routes of adaptation.

CNVs are frequently encountered within the timescales of ALE experiments, suggesting it as an effective mechanism of short-scale adaptation[13]. Interesting within this study however was varying frequency and type of CNVs within the evolved strains. ALEX-1, for instance, was able to outperform ALEX-2 in terms of biomass production and xylose consumption despite the latter having a far greater number of contigs with CNVs than the former (5 vs 38). In a similar way, the performance of ALEX-1, -2, -3 and -4, although improved relative to the progenitor, was far behind ALEX-5 despite the overall number of CNVs for this strain not standing out (though the fact they were all duplications may be a contributing factor). These results suggest a situation where there are different evolutionary routes to improved xylose metabolism, each with different degrees of phenotypic success. As discussed, ALEX-5 was the only strain to have a CNV for a gene within the xylose metabolic pathway, *XKS1,* but was in common with the other evolved strains in having reduced copy number of some *GAL* genes. A study which evaluated the routes to improved copper tolerance through ALE found a similar result, in that the majority of strains (27/34) had increased copy number of the *CUP1* gene, but individual strains acquired further tolerance through other routes, such as aneuploidy of chromosomes II or VIII or SNPs to resistance-related genes[45]. It is conceivable that given more time, more strains would have joined ALEX-5 with the *XKS1* CNV, unless other CNVs were somehow blocking the development of such CNVs elsewhere.

A second interesting result from the whole genome sequencing was that absolute frequency of CNVs does not correlate with adapted xylose performance. Rather, it appears that the selective pressure applied in this instance was conducive in allowing strains to accumulate CNVs at no detriment to survival. Within this hypothesis, the high number of CNVs found could instead be a consequence of there being only a narrow range of genes capable of improving xylose metabolism (such as *XKS1*), meaning cells were 'fishing' for beneficial changes by indiscriminately changing the copy number without any negative selection. This was most likely facilitated as growth on xylose provides a non-lethal selective pressure, where the detrimental effects of CNVs were otherwise masked. This situation would be unlikely to present itself where a selective pressure is potentially lethal, and induces a stress response central to many essential pathways, such as a cocktail of fermentation inhibitors in the presence of glucose, for example.

6.4.6. SNP variants within ALEX strains.

Following CNV analysis, the SnpEff tool was used to screen evolved strain sequence data for single nucleotide polymorphisms[21]. Due to the number of strains sequenced, only high impact SnpEffs were considered for this analysis. Within this criteria are several putative impacts including the loss of a start or stop codon, splice site mutation or frameshift mutations. Throughout the strains there were 21 high impact SNP variants, including three SNPs within the start codon for *AIM18*, *TAM41* and *THP3/VMS1* (Table 3). Contrary to CNV data (Table 1), ALEX-2 had both the fewest total and unique high impact SNPs, while the other strains shared a similar frequency. *JLP2* was the only variant found across all strains, though 5 other SNP variants were shared across four strains. Jlp2 is a member of the J-protein family, which function as chaperone cofactors to Hsp70 heat shock proteins[46]. Given the limited literature of J-proteins, and *JLP2* in particular, it is difficult to suggest a reason why selection within xylose media has caused all strains to independently develop a variant within this gene. Among the other shared high impact SNPs, *SIP5*, which is known to interact with *SNF1,* may have a more obvious potential phenotypic effect

as the two proteins interact in response to glucose limitation, facilitating alternate carbon utilisation[47]. It is not clear from the data available how the SNP is affecting gene function, however it is significant that this variant was found in all strains except ALEX-2, which had the lowest biomass production from xylose of the evolved strains. *ITR2*, another commonly shared SNP variant, encodes a myo-inositol transporter protein which has sequence homology to other sugar transporters including the *E. coli* arabinose and xylose transporters (xylE and araE), a *Kluveromyces lactis* lactose transporter (*LAC12)* and a *S. cerevisiae* maltose (*MAL61*), galactose (*GAL2*) and glucose (*SNF3*) transporter[48]. *YAK1,* which also appears in four out of five strains as a high impact variant, has a crucial role in the expression of glucose-repressed genes by phosphorylating the Pop2 transcription factor[49]. Transcription of *YAK1* and subsequent phosphorylation of Pop2 increases during stationary phase and correlates with cell cycle arrest. Pop2 mutants, a phenotype which could be mimicked by loss of Yak1 function, had increased rates of re-replication[50]. A hypothesis therefore is that high-impact SNPs within *YAK1* could lead to high rates of genome instability causing gene amplification, aneuploidy and polyploidy; structural variants found within all ALEX evolved strains[51]. Interestingly however, this SnpEff was not found within the strain with the greatest number of CNV contigs, ALEX-2. This suggests that multiple mechanisms leading to CNVs are at play. In fact, Mms22 knockouts, a protein which carried a SNP in ALEX-4, have increased frequency of double stranded breaks, with knockout strains having delayed clearance of broken chromosomes compared with wildtype cells[52].

ALEX-5 unique SNPs do not appear directly linked to any observed metabolic phenotype, but they do appear to be crucial with respect to maintenance of DNA structure. *SCP160*, a gene necessary for maintenance of exact ploidy, was found with a high-impact variant[53,54]. Increased ploidy (to 4X coverage) was an ALEX-5 specific structural variant, with all other strains possessing aneuploid CNVs. *TTI1* also plays a role in maintaining DNA quality, as it encodes a protein which interacts with telomere binding protein Tel2[55]. *TEL2* itself is an essential gene regulating telomere length within *S. cerevisiae,* and disrupting a copy of this gene led to decreased teleomere length within mutant strains[56]. The loss of telomeres within yeast results in

telomere shortening, activation of DNA damage checkpoints and cell death. It's been described that cells able to survive telomere loss have instead had telomeres maintained via recombination, which could potentially introduce ploidy changes[57]. Aside from this, ALEX-5 had one more unique SNPs within *STI1*, which like *JLP2*, encodes a heat shock protein cochaperone[58].

Table 3. High impact SNP variants as determined by SnpEff. The *S. cerevisiae* homologue for each *M. pulcherrima* gene is shown. In instances where protein sequence homology was >95% for two genes, both are listed. 'St Co' refers to a SNP within the start codon.

SC Gene	ALEX-1	ALEX-2	ALEX-3	ALEX-4	ALEX-5	Freq.
JLP2						5
FRA1						4
SIP5						4
ECM32						4
ITR2						4
YAK1						4
THP3/VMS1						2
St Co*: THP3/VMS1*						2
ELA1/RGL1						2
PPN1						1
BPL1						1
FLO9						1
INP51						1
AIM18						1
St Co: *AIM18*						1
MMS22						1
TAM41						1
St Co: *TAM41*						1
STI1						1
SCP160						1
TTI1						1
Total	12	3	9	8	11	
Unique	3	1	2	3	3	

6.5. Conclusion

Starting with five parallel cell lines, serial batch culturing in media containing xylose as the sole carbon source produced five final strains capable of producing greater biomass and consumption of xylose than the progenitor strain. Though all final strains had a significant improvement, the degree of this improvement, as well as the severity of evolutionary trade-offs with respect to lipid production and glucose growth, suggests different routes to xylose adaptation. It was unsurprising therefore that a large variation between evolved strains was seen after performing whole genome sequencing, particularly with respect to the frequency of CNVs. Interestingly, a correlation was not seen between frequency of contigs with CNVs and improved xylose metabolism, indicating that a lot of this variation was having no advantage, or disadvantage, to overall fitness relative to the population. It is hypothesised therefore that xylose, as a non-lethal selective pressure and a carbon source with a relatively narrow metabolic pathway, is able to provide an evolutionary environment where potential deleterious CNVs are allowed to accumulate as long as they do not negatively affect xylose metabolism. This is supported by the respective performance of ALEX-1, which contained relatively few CNVs but few evolutionary trade-offs and ALEX-2, which had a significantly higher frequency of CNV but performed comparatively poorly on glucose and lipid production. The evolutionary trade-offs in the final strains highlight the unpredictability of ALE as a method of strain improvement. That said, this study has highlighted several potential targets to improve xylose metabolism when a reverse engineering approach is considered.

6.6. References

1. Reijnders, L.; Huijbregts, M.A.J. Palm oil and the emission of carbon-based greenhouse gases. *J. Clean. Prod.* 2008.
2. Long, Z.; Santomauro, F.; Fan, J.; Macquarrie, D.; Clark, J.; Chuck, C.J.; Budarin, V. Fast microwave-assisted acidolysis: a new biorefinery approach for the zero-waste utilisation of lignocellulosic biomass to produce high quality lignin and fermentable saccharides. *Faraday Discuss.* 2017.
3. Sànchez Nogué, V.; Karhumaa, K. Xylose fermentation as a challenge for commercialization of lignocellulosic fuels and chemicals. *Biotechnol. Lett.* 2015.
4. Santamauro, F.; Whiffin, F.M.; Scott, R.J.; Chuck, C.J. Low-cost lipid production by an oleaginous yeast cultured in non-sterile conditions using model waste resources. *Biotechnol. Biofuels* 2014, *7*, 34.
5. Scalcinati, G.; Otero, J.M.; Van Vleet, J.R.H.; Jeffries, T.W.; Olsson, L.; Nielsen, J. Evolutionary engineering of Saccharomyces cerevisiae for efficient aerobic xylose consumption. *FEMS Yeast Res.* 2012.
6. Sonderegger, M.; Sauer, U. Evolutionary engineering of Saccharomyces cerevisiae for anaerobic growth on xylose. *Appl. Environ. Microbiol.* 2003.
7. Becker, J.; Boles, E. A modified Saccharomyces cerevisiae strain that consumes L-arabinose and produces ethanol. *Appl. Environ. Microbiol.* 2003.
8. Ho, P.W.; Swinnen, S.; Duitama, J.; Nevoigt, E. The sole introduction of two single-point mutations establishes glycerol utilization in Saccharomyces cerevisiae CEN.PK derivatives. *Biotechnol. Biofuels* 2017.
9. Strucko, T.; Zirngibl, K.; Pereira, F.; Kafkia, E.; Mohamed, E.T.; Rettel, M.; Stein, F.; Feist, A.M.; Jouhten, P.; Patil, K.R.; et al. Laboratory evolution reveals regulatory and metabolic trade-offs of glycerol utilization in Saccharomyces cerevisiae. *Metab. Eng.* 2018.
10. Wouter Wisselink, H.; Toirkens, M.J.; Wu, Q.; Pronk, J.T.; Van Maris, A.J.A. Novel evolutionary engineering approach for accelerated utilization of glucose, xylose, and arabinose mixtures by engineered Saccharomyces cerevisiae strains. *Appl. Environ. Microbiol.* 2009.
11. Farwick, A.; Bruder, S.; Schadeweg, V.; Oreb, M.; Boles, E. Engineering of yeast hexose transporters to transport D-xylose without inhibition by D-glucose. *Proc. Natl. Acad. Sci. U. S. A.* 2014, *111*, 5159–64.
12. Guan, Y.; Dunham, M.J.; Troyanskaya, O.G. Functional analysis of gene duplications in Saccharomyces cerevisiae. *Genetics* 2007, *175*, 933–943.
13. Gorter de Vries, A.R.; Pronk, J.T.; Daran, J.M.G. Industrial relevance of chromosomal copy number variation in Saccharomyces yeasts. *Appl. Environ. Microbiol.* 2017.
14. Hicks, R.H.; Sze, Y.; Chuck, C.J.; Henk, D.A. Enhanced inhibitor tolerance and increased lipid productivity through adaptive laboratory evolution in the oleaginous yeast Metshnikowia pulcherrima. Manuscript submitted for publication. 2018.
15. RStudio RStudio: Integrated Development for R 2015.
16. Bligh, E.G.; Dyer, W.J. A rapid method of total lipid extraction and purification. *Can. J. Biochem. Physiol.* 1959, *37*, 911–917.

17. Safonova, Y.; Bankevich, A.; Pevzner, P.A. DIPSPADES: Assembler for highly polymorphic diploid genomes. In Proceedings of the Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); 2014.
18. Stanke, M.; Morgenstern, B. AUGUSTUS: A web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* 2005.
19. Byrne, K.P.; Wolfe, K.H. The Yeast Gene Order Browser: Combining curated homology and syntenic context reveals gene fate in polyploid species. *Genome Res.* 2005.
20. Li, H.; Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 2010.
21. Cingolani, P.; Platts, A.; Wang, L.L.; Coon, M.; Nguyen, T.; Wang, L.; Land, S.J.; Lu, X.; Ruden, D.M. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly (Austin).* 2012.
22. Mans, R.; Daran, J.M.G.; Pronk, J.T. Under pressure: evolutionary engineering of yeast strains for improved performance in fuels and chemicals production. *Curr. Opin. Biotechnol.* 2018.
23. Opulente, D.A.; Morales, C.M.; Carey, L.B.; Rest, J.S. Coevolution Trumps Pleiotropy: Carbon Assimilation Traits Are Independent of Metabolic Network Structure in Budding Yeast. *PLoS One* 2013.
24. Sitepu, I.; Selby, T.; Lin, T.; Zhu, S.; Boundy-Mills, K. Carbon source utilization and inhibitor tolerance of 45 oleaginous yeast species. *J. Ind. Microbiol. Biotechnol.* 2014, *41*, 1061–1070.
25. Dhar, R.; Sägesser, R.; Weikert, C.; Wagner, A. Yeast adapts to a changing stressful environment by evolving cross-protection and anticipatory gene regulation. *Mol. Biol. Evol.* 2013.
26. Dragosits, M.; Mozhayskiy, V.; Quinones-Soto, S.; Park, J.; Tagkopoulos, I. Evolutionary potential, cross-stress behavior and the genetic basis of acquired stress resistance in Escherichia coli. *Mol. Syst. Biol.* 2013.
27. Berry, D.B.; Guan, Q.; Hose, J.; Haroon, S.; Gebbia, M.; Heisler, L.E.; Nislow, C.; Giaever, G.; Gasch, A.P. Multiple means to the same end: The genetic basis of acquired stress resistance in yeast. *PLoS Genet.* 2011.
28. Jeffries, T.W. Engineering yeasts for xylose metabolism. *Curr. Opin. Biotechnol.* 2006.
29. Bettiga, M.; Bengtsson, O.; Hahn-H??gerdal, B.; Gorwa-Grauslund, M.F. Arabinose and xylose fermentation by recombinant Saccharomyces cerevisiae expressing a fungal pentose utilization pathway. *Microb. Cell Fact.* 2009.
30. Chen, G.; Rubinstein, B.; Li, R. Whole chromosome aneuploidy: Big mutations drive adaptation by phenotypic leap. *BioEssays* 2012.
31. Steensels, J.; Snoek, T.; Meersman, E.; Nicolino, M.P.; Voordeckers, K.; Verstrepen, K.J. Improving industrial yeast strains: exploiting natural and artificial diversity. *FEMS Microbiol. Rev.* 2014, 1–49.
32. Slot, J.C.; Rokas, A. Multiple GAL pathway gene clusters evolved independently and by different mechanisms in fungi. *Proc. Natl. Acad. Sci.* 2010.
33. Bisson, L.F.; Coons, D.M.; Kruckeberg, A.L.; Lewis, D.A. Yeast sugar

transporters. *Crit. Rev. Biochem. Mol. Biol.* 1993.
34. Egriboz, O.; Jiang, F.; Hopper, J.E. Rapid GAL gene switch of Saccharomyces cerevisiae depends on nuclear Gal3, not nucleocytoplasmic trafficking of Gal3 and Gal80. *Genetics* 2011.
35. Douglas, H.C.; Hawthorne, D.C. Enzymatic Expression and Genetic Linkage of Genes Controlling Galactose Utilization in Saccharomyces. *Genetics* 1964.
36. Majumdar, S.; Ghatak, J.; Mukherji, S.; Bhattacharjee, H.; Bhaduri, A. UDPgalactose 4-epimerase from Saccharomyces cerevisiae: A bifunctional enzyme with aldose 1-epimerase activity. *Eur. J. Biochem.* 2004.
37. Steenwyk, J.L.; Rokas, A. Copy number variation in fungi and its implications for wine yeast genetic diversity and adaptation. *Front. Microbiol.* 2018.
38. Yang, F.; Zhang, S.; Zhou, Y.J.; Zhu, Z.; Lin, X.; Zhao, Z.K. Characterization of the mitochondrial NAD +-dependent isocitrate dehydrogenase of the oleaginous yeast Rhodosporidium toruloides. *Appl. Microbiol. Biotechnol.* 2012.
39. Ames, R.M.; Rash, B.M.; Hentges, K.E.; Robertson, D.L.; Delneri, D.; Lovell, S.C. Gene duplication and environmental adaptation within yeast populations. *Genome Biol. Evol.* 2010.
40. Rodriguez-Peña, J.M.; Cid, V.J.; Arroyo, J.; Nombela, C. The YGR194c (XKS1) gene encodes the xylulokinase from the budding yeast Saccharomyces cerevisiae. *FEMS Microbiol. Lett.* 1998.
41. Toivari, M.H.; Aristidou, A.; Ruohonen, L.; Penttilä, M. Conversion of xylose to ethanol by recombinant Saccharomyces cerevisiae: Importance of xylulokinase (XKS1) and oxygen availability. *Metab. Eng.* 2001.
42. Sinha, A.; Maitra, P.K. Induction of specific enzymes of the oxidative pentose phosphate pathway by glucono-delta-lactone in Saccharomyces cerevisiae. *J. Gen. Microbiol.* 1992.
43. Kamisaka, Y.; Noda, N.; Tomita, N.; Kimura, K.; Kodaki, T.; Hosaka, K. Identification of genes affecting lipid content using transposon mutagenesis in Saccharomyces cerevisiae. *Biosci. Biotechnol. Biochem.* 2006.
44. Zhou, Y.J.; Buijs, N.A.; Zhu, Z.; Qin, J.; Siewers, V.; Nielsen, J. Production of fatty acid-derived oleochemicals and biofuels by synthetic yeast cell factories. *Nat. Commun.* 2016.
45. Gerstein, A.C.; Ono, J.; Lo, D.S.; Campbell, M.L.; Kuzmin, A.; Otto, S.P. Too much of a good thing: The unique and repeated paths toward copper adaptation. *Genetics* 2014.
46. Walsh, P.; Bursać, D.; Law, Y.C.; Cyr, D.; Lithgow, T. The J-protein family: Modulating protein assembly, disassembly and translocation. *EMBO Rep.* 2004.
47. Sanz, P.; Ludin, K.; Carlson, M. Sip5 interacts with both the Reg1/Glc7 protein phosphatase and the Snf1 protein kinase of Saccharomyces cerevisiae. *Genetics* 2000.
48. Nikawa, J.; Tsukagoshi, Y.; Yamashita, S. Isolation and characterization of two distinct myo-inositol transporter genes of Saccharomyces cerevisiae. *J. Biol. Chem.* 1991.
49. Moriya, H.; Shimizu-Yoshida, Y.; Omori, A.; Iwashita, S.; Katoh, M.; Sakai, A. Yak1p, a DYRK family kinase, translocates to the nucleus and phosphorylates

yeast Pop2p in response to a glucose signal. *Genes Dev.* 2001.
50. Toda, T.; Ochotorena, I.; Kominami, K.I. Two distinct ubiquitin-proteolysis pathways in the fission yeast cell cycle. *Philos. Trans. R. Soc. B Biol. Sci.* 1999.
51. Truong, L.N.; Wu, X. Prevention of DNA re-replication in eukaryotic cells. *J. Mol. Cell Biol.* 2011.
52. Ben-Aroya, S.; Agmon, N.; Yuen, K.; Kwok, T.; McManus, K.; Kupiec, M.; Hieter, P. Proteasome nuclear activity affects chromosome stability by controlling the turnover of Mms22, a protein important for DNA repair. *PLoS Genet.* 2010.
53. Wintersberger, U.; Kühne, C.; Karwan, A. Scp160p, a new yeast protein associated with the nuclear membrane and the endoplasmic reticulum, is necessary for maintenance of exact ploidy. *Yeast* 1995.
54. Marsellach, F.X.; Huertas, D.; Azorín, F. The multi-KH domain protein of Saccharomyces cerevisiae Scp160p contributes to the regulation of telomeric silencing. *J. Biol. Chem.* 2006.
55. Shevchenko, A.; Roguev, A.; Schaft, D.; Buchanan, L.; Habermann, B.; Sakalar, C.; Thomas, H.; Krogan, N.J.; Shevchenko, A.; Stewart, A.F. Chromatin Central: Towards the comparative proteome by accurate mapping of the yeast proteomic environment. *Genome Biol.* 2008.
56. Runge, K.W.; Zakian, V.A. TEL2, an essential gene required for telomere length regulation and telomere position effect in Saccharomyces cerevisiae. *Mol. Cell. Biol.* 1996.
57. Millet, C.; Ausiannikava, D.; Le Bihan, T.; Granneman, S.; Makovets, S. Cell populations can use aneuploidy to survive telomerase insufficiency. *Nat. Commun.* 2015.
58. Reidy, M.; Kumar, S.; Anderson, D.E.; Masison, D.C. Dual roles for yeast Sti1/Hop in regulating the Hsp90 chaperone cycle. *Genetics* 2018.

Table S1. **Summary of gene coverage within evolved strains.** Values refer to contigs which had 2x coverage (2 copies) within the progenitor but are now represented with either 3X (blue) or 4X (gold) copies. Frequency of the contig modifications and unique-to-strain CNVs are tallied.

Contig	ALEX-1	ALEX-2	ALEX-3	ALEX-4	ALEX-5	Freq.
71						4
95						4
100						4
111						4
122						4
124						4
127						4
101						3
104						3
144						3
10						2
13						2
18						2
41						2
42						2
46						2
57						2
5						1
9						1
11						1
14						1
17						1
20						1
23						1
24						1
25						1
26						1
27						1
34						1
36						1
40						1
43						1
44						1
45						1
47						1
48						1
50						1
54						1
59						1
76						1
78						1
80						1
91						1
97						1
98						1
99						1
110						1
113						1
116						1
117						1
121						1
123						1
133						1
136						1
141						1
145						1
146						1
Unique CNVs	1	24	1	4	10	

This declaration concerns the article entitled:										
The oleaginous yeast *M. pulcherrima* improves xylose utilisation through different routes during adaptive laboratory evolution										
Publication status (tick one)										
draft manuscript	**X**	**Submitted**		**In review**		**Accepted**		**Published**		
Publication details (reference)										
Candidate's contribution to the paper (detailed, and also given as a percentage).	The candidate contributed to/ considerably contributed to/predominantly executed the… Formulation of ideas: Directed evolution idea was formulated by myself. (100%) Design of methodology: Methodology of 'wet' experimental work was designed by myself. Dr. Daniel Henk guided and assisted work based upon whole genome sequencing. (75%) Experimental work: The first half of batch culturing was performed by Jenny Thomas under my supervision, the second half was performed by myself. All 'wet' characterisation experiments in follow up were performed by myself. HPLC was ran by Felix Abeln. Dr. Daniel Henk provided guidance and assistance in transforming read data from whole genome sequencing and analysis of this data was performed by myself. Presentation of data in journal format: Applied Microbiology and Biotechnology									
Statement from Candidate	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.									
Signed						**Date**				

7. Conclusions and Further Work.

Work within this thesis was split into two major workflows; the development of a reliable high throughput lipid estimation method, and the use of ALE to develop strains of *M. pulcherrima* with improved industrial phenotypes. In the first ALE strategy, a starting strain was successfully adapted to either a cocktail of fermentation inhibitors, or a single inhibitor, formic acid. Though both strategies were successful in their aim of improving inhibitor tolerance, strains adapted to the inhibitor cocktail now produced almost double the amount oil compared with the progenitor. In the second round of ALE, a candidate strain from the fermentation inhibitor cocktail strategy, 4x3, was grown sequentially in media containing xylose as the sole carbon source. All resulting evolved strains now possessed greater utilisation of xylose, including one strain, ALEX-5, which now produced biomass over 17 times greater than the progenitor when xylose was the sole carbon source.

Due to the unexpected result of increased lipid production within strains adapted to a fermentation inhibitor cocktail, it was possible to take forward the candidate strain 4x3 for screening at large scale. This screening was improved through the use of the lipid estimation method, which also provides data on culture population dynamics. This allowed information to be gathered about the suitability of this strain to be grown semi-continuously at large scales and to identify and overcome potential production bottlenecks. Given this, follow up work from this thesis would be mainly split into:

a) Understanding, testing and confirming the genetic basis leading to the observed phenotypes within evolved strains of *M. pulcherrima*,
b) Continuing scale up process understanding and utilise biological and engineering routes to optimise productivity and product formation.

Genetic basis of strain improvement – Specific targets to study.

Performing whole genome sequencing revealed a host of genes with altered copy number within the evolved strains which may contribute towards the acquired phenotypes. By working under the assumption that copy number is directly correlated with expression level, overexpressing these candidate genes may help determine causality. An obvious example would be to overexpress the xylose metabolism related gene, *XKS1*[143], which was only found with increased copy number in ALEX-5; the strain with substantially improved biomass production on xylose compared with the progenitor and other evolved cell lines. For the strains with improved fermentation inhibitor tolerance, overexpression targets include *MSN2* and *PDR15* which have both been shown to generate strains with improved tolerance when overexpressed in *S. cerevisiae*[126,128].

An important phenotype occurring throughout this study was improved lipid production in strains evolved in the presence of an inhibitor cocktail. Due to this being an unexpected phenotype, it was anticipated that the factors causing increased lipid production were as a side effect of improved tolerance to furfural and 5-HMF, as production from strains adapted to formic acid reduced. A working theory was therefore proposed that adaptations leading to increased cellular levels of NADPH, a co factor essential for enzymes detoxifying aldehyde inhibitors, as well as a key building block in lipid biosynthesis, may have occurred within these strains. This hypothesis was supported due to the fact that oleaginous organisms are found with higher cellular levels than non-oleaginous, although the source of this NADPH varies between species[23]. It was therefore interesting to find that *STB5,* a gene which positively regulates NADPH producing enzymes including *ALD6*, a gene involved in aldehyde tolerance, was found with increased copy number[51,144]. Overexpression studies of this gene could contribute towards understanding of several key questions: Does increased expression lead to increased cellular NADPH levels? Does overexpression of *STB5* alone lead to increased fermentation inhibitor tolerance? Does increased NADPH levels lead to increased lipid accumulation in *M. pulcherrima*? An interesting follow on experiment from this would be to overexpress *ALD6* itself, or knocking out *ADH6,* which also plays a role in aldehyde tolerance but consumes rather than replenishes NADPH[136].

At present, 4x3 is the only strain which has been analysed from the strains adapted to a fermentation inhibitor cocktail. Performing whole genome sequencing on the other three strains, 4x1, 4x2 and 4x4, and looking for shared genetic changes will help to better elucidate what genetic changes are potentially causative, given the phenotypic similarly between these strains. Sequencing of the strains adapted to just formic acid will also be required. By analysing these strains, a comparison of the potential routes taken by 4x3 to improve tolerance to acid stress when present within an inhibitor cocktail, versus the adaptive routes of strains adapted only to formic acid can be made. In addition, adaptation to formic acid caused the reverse effect on lipid content. This supports the theory that adaption to aldehyde inhibitors is causative in increased lipid production, but also poses an interesting question of why lipid production reduced.

Genetic basis of strain improvement – Routes of adaptation.

An interesting result from the second ALE experiment to improve xylose metabolism was the variation between the five evolved cell lines in their frequency of CNVs to large sections of the genome, and yet there being no correlation between number of CNV and degree of improved metabolism. In fact, the evolved strain with the highest number of CNVs improved the least relative to the other strains. These results suggest in one instance that adaption though CNVs is an effective route for strains, but that many of the CNVs were redundant in relation to improved metabolism of xylose. The reason, it is hypothesised, was that under the non-lethal selective pressure of xylose, strains were able to frequently change their genetic structure in search of gaining a CNV related to one of the few xylose related genes, leading to CNVs of many unrelated genes.

It would be interesting therefore to compare the abundance of CNV occurring in strains adapted to a non-lethal stress, such as xylose, versus those adapted to a potentially lethal stress, such as fermentation inhibitors. Whilst work in this thesis did complete two such ALE studies, given that the strain used in the xylose study,

4x3, came from a previous ALE background, it is not an ideal condition to make such a comparison. What is potentially significant however, is comparing the phenotypic similarity of the final strains at the end of the three ALE strategies. For instance, when NCYC2580 was adapted to a cocktail of inhibitors or formic acid, all final strains from both strategies shared a very similar growth and lipid phenotypes respectively. In comparison, variation amongst the final xylose evolved strains was vast, for instance lipid production ranged from 31.8% for the highest strain, to 8.7% for the lowest. The phenotypic results alone therefore suggest that adaptation to potentially lethal selective pressures have to follow a common route, where as non-lethal selection allows for multiple routes to occur, giving rise to the multiple phenotypes observed in just five cell lines.

With this in mind, it would also be interesting to re-approach adaptation to xylose but using a dynamic evolution strategy whereby the evolving cell population was intermittently passed through media containing glucose, rather than continuously culturing in only xylose. Strategies such as this have been shown to reduce evolutionary trade-offs, a factor prevalent in all xylose-evolved strains as shown by significantly reduced biomass from glucose compared with the progenitor[91]. Introducing an intermittent glucose batch within the evolution process could potentially have the same effect as a lethal selective pressure, whereby strains which have acquired CNVs which improve xylose metabolism but also accumulated CNVs deleterious to glucose metabolism, would be lost from the population. It is possible in this instance that given growth on glucose implicates a far greater number of genes than does xylose, only CNVs directly affecting xylose, such as *XKS1*, may survive such a selection regime. This could in turn lead to phenotypically similar final strains as potential routes to adaptation are narrowed.

Understanding and optimising a scale up process.

Through the use of cell size analysis, culture population dynamics can be monitored and lipid accumulation estimated. This method supported an initial study which evaluated the suitability of *M. pulcherrima* to a semi continuous process, where a

potential production bottleneck was identified which caused the oleaginous phenotype to be lost when cultured in successive batches. This bottleneck was overcome by seeding a small quantity of new cells from an SMB overnight into the main culture when there was a harvesting of the main culture and addition of fresh media. With this strategy, different semi continuous models were tested at small scale, which altered the degree of media removal/fresh media addition at the end of each batch. Here, it was found that increasing the percentage of existing culture being removed at each batch (with 50% removal the highest trialled) ensured that 'older' cells, with the potential to switch to a non-oleaginous cell type, were being taken out of the culture before this was allowed to occur. Importantly, the fresh cells seeded in at the same point were fast to grow and accumulate oil, meaning overall biomass did not reduce. The understanding from the small scale work was then validated at large scale at 250 L. Here, although the small scale models were not followed rigidly for logistical reasons, promising results were obtained showing that biomass and lipid production was maintained at an almost consistent level for 16 days when culture was regularly harvested and replaced.

With these experiments establishing that a semi continuous culture cannot be sustained unless fresh cells are seeded into the main culture, further optimisation is now required. There are several key parameters to investigate, namely: what is the lowest possible quantity of fresh cells to seed into the main culture which provides consistency in lipid accumulation and biomass of the overall culture, and what ratio of culture removal/media addition and how frequent this occurs for optimal biomass and lipid production to be maintained. Finding the lowest quantity of fresh cells to add to the main culture is important in reducing potential costs at large scale, as it appears this is an unavoidable measure.

Another interesting opportunity to investigate from a processing perspective is how a semi continuous model can be tailored to affect the properties of the oil. Work conducted within bioreactors using 4x3 showed that in the first 24 hours, the fatty acid profile was approximately 50% C16:0, 50% C18:1, with this shifting to 35% C16:0, 55% C18:1 and C16:1 and C18:2 0% after 72 hours. To obtain an oil similar to palm

oil, maintaining a high proportion of C16:0 is essential, with palm oil itself consisting of 44% C16:0 palmitic acid. It is important therefore that investigation into an optimised semi continuous process considers fatty acid composition. For instance, if a 1/3rd of the culture is harvested every three days, then some cells could remain for greater than six days; a point at which C16:0 levels drops to 30%. There is likely therefore to be a trade off between allowing sufficient time for lipid accumulation, and not allowing cells to remain for long enough for ratios of C16:0 to drop. It is also unknown at this point whether the ratio of C16:0 to C18:1 changes due to the former being converted to the latter, or whether production of C16:0 stops within an early growth phase, leaving C18:1 fatty acids to continue accumulating and predominating the FA profile.

By evaluating and utilising culturing strategies such as these, it lessens the requirement for genetic engineering to produce the desired fatty acid composition. That said, successfully targeting fatty acid desaturase genes in knockout studies has the potential to shift the compromise between time given for lipid accumulation and fatty acid composition.

8. References

1. Basiron, Y. Palm oil production through sustainable plantations. *Eur. J. Lipid Sci. Technol.* 2007.
2. Vijay, V.; Pimm, S.L.; Jenkins, C.N.; Smith, S.J. The Impacts of Oil Palm on Recent Deforestation and Biodiversity Loss. *PLoS One* 2016.
3. Petrenko, C.; Paltseva, J.; Searle, S. *ECOLOGICAL IMPACTS OF PALM OIL EXPANSION IN INDONESIA*; 2016;
4. Fitzherbert, E.B.; Struebig, M.J.; Morel, A.; Danielsen, F.; Brühl, C.A.; Donald, P.F.; Phalan, B. How will oil palm expansion affect biodiversity? *Trends Ecol. Evol.* 2008.
5. Malins, C. *For peat's sake: Understanding the climate implications of palm oil biodiesel consumption*; 2017;
6. Environment, T.& *Globiom: the basis for biofuel policy post-2020*; 2016;
7. Malins, C. *Driving deforestation: The impact of expanding palm oil demand through biofuel policy*; 2018;
8. Kharina, A.; Malins, C.; Searle, S. *BIOFUELS POLICY IN INDONESIA: OVERVIEW AND STATUS REPORT*; 2016;
9. on palm oil and deforestation of rainforests.
10. Palm Oil: Outcome of the Trilogue of the EU's Renewable Energy Directive (RED II) Available online: https://eeas.europa.eu/delegations/indonesia/46646/palm-oil-outcome-trilogue-eu's-renewable-energy-directive-red-ii_it.
11. Transport & Environment Seven facts about palm oil biodiesel Available online: https://www.transportenvironment.org/publications/seven-facts-about-palm-oil-biodiesel (accessed on Oct 10, 2018).
12. Chisti, Y. Biodiesel from microalgae beats bioethanol. *Trends Biotechnol.* 2008, *26*, 126–131.
13. Sitepu, I.R.; Garay, L.A.; Sestric, R.; Levin, D.; Block, D.E.; German, J.B.; Boundy-Mills, K.L. Oleaginous yeasts for biodiesel: Current and future trends in biology and production. *Biotechnol. Adv.* 2014, *32*, 1336–1360.
14. Ageitos, J.M.; Vallejo, J.A.; Veiga-Crespo, P.; Villa, T.G. Oily yeasts as oleaginous cell factories. *Appl. Microbiol. Biotechnol.* 2011, *90*, 1219–1227.
15. Santamauro, F.; Whiffin, F.M.; Scott, R.J.; Chuck, C.J. Low-cost lipid production by an oleaginous yeast cultured in non-sterile conditions using model waste resources. *Biotechnol. Biofuels* 2014, *7*, 34.
16. Leiva-Candia, D.E.; Pinzi, S.; Redel-Macías, M.D.; Koutinas, A.; Webb, C.; Dorado, M.P. The potential for agro-industrial waste utilization using oleaginous yeast for the production of biodiesel. *Fuel* 2014, *123*, 33–42.
17. Valle-Rodríguez, J.O.; Shi, S.; Siewers, V.; Nielsen, J. Metabolic engineering of Saccharomyces cerevisiae for production of fatty acid ethyl esters, an advanced biofuel, by eliminating non-essential fatty acid utilization pathways. *Appl. Energy* 2014, *115*, 226–232.
18. Easterling, E.R.; French, W.T.; Hernandez, R.; Licha, M. The effect of glycerol as a sole and secondary substrate on the growth and fatty acid composition of Rhodotorula glutinis. *Bioresour. Technol.* 2009, *100*, 356–361.
19. Fakas, S. Lipid biosynthesis in yeasts: A comparison of the lipid biosynthetic

pathway between the model nonoleaginous yeast Saccharomyces cerevisiae and the model oleaginous yeast Yarrowia lipolytica. *Eng. Life Sci.* 2017.
20. Tai, M.; Stephanopoulos, G. Engineering the push and pull of lipid biosynthesis in oleaginous yeast Yarrowia lipolytica for biofuel production. *Metab. Eng.* 2013.
21. Hiesinger, M.; Wagner, C.; Schüller, H.J. The acetyl-CoA synthetase gene ACS2 of the yeast: Saccharomyces cerevisiae is coregulated with structural genes of fatty acid biosynthesis by the transcriptional activators Ino2p and Ino4p. *FEBS Lett.* 1997.
22. Vorapreeda, T.; Thammarongtham, C.; Cheevadhanarak, S.; Laoteng, K. Alternative routes of acetyl-CoA synthesis identified by comparative genomic analysis: Involvement in the lipid production of oleaginous yeast and fungi. *Microbiology* 2012.
23. Wasylenko, T.M.; Ahn, W.S.; Stephanopoulos, G. The oxidative pentose phosphate pathway is the primary source of NADPH for lipid overproduction from glucose in Yarrowia lipolytica. *Metab. Eng.* 2015.
24. Zhang, Y.; Adams, I.P.; Ratledge, C. Malic enzyme: The controlling activity for lipid production? Overexpression of malic enzyme in Mucor circinelloides leads to a 2.5-fold increase in lipid accumulation. *Microbiology* 2007, *153*, 2013–2025.
25. Morin, N.; Cescut, J.; Beopoulos, A.; Lelandais, G.; Le Berre, V.; Uribelarrea, J.L.; Molina-Jouve, C.; Nicaud, J.M. Transcriptomic analyses during the transition from biomass production to lipid accumulation in the oleaginous yeast Yarrowia lipolytica. *PLoS One* 2011.
26. Ledesma-Amaro, R.; Nicaud, J.M. Yarrowia lipolytica as a biotechnological chassis to produce usual and unusual fatty acids. *Prog. Lipid Res.* 2016, *61*, 40–50.
27. Abghari, A.; Chen, S. Yarrowia lipolytica as an Oleaginous Cell Factory Platform for Production of Fatty Acid-Based Biofuel and Bioproducts. *Front. Energy Res.* 2014.
28. Koutinas, A.A.; Chatzifragkou, A.; Kopsahelis, N.; Papanikolaou, S.; Kookos, I.K. Design and techno-economic evaluation of microbial oil production as a renewable resource for biodiesel and oleochemical production. *Fuel* 2014, *116*, 566–577.
29. Johari, A.; Nyakuma, B.B.; Mohd Nor, S.H.; Mat, R.; Hashim, H.; Ahmad, A.; Yamani Zakaria, Z.; Tuan Abdullah, T.A. The challenges and prospects of palm oil based biodiesel in Malaysia. *Energy* 2015.
30. Balan, V. Current Challenges in Commercially Producing Biofuels from Lignocellulosic Biomass. *ISRN Biotechnol.* 2014.
31. Mathews, J.A.; Tan, H.; Moore, M.J.B.; Bell, G. A conceptual lignocellulosic “feed+fuel” biorefinery and its application to the linked biofuel and cattle raising industries in Brazil. *Energy Policy* 2011.
32. Ho, D.P.; Ngo, H.H.; Guo, W. A mini review on renewable sources for biofuel. *Bioresour. Technol.* 2014.
33. Gomez, L.D.; Steele-King, C.G.; McQueen-Mason, S.J. Sustainable liquid biofuels from biomass: The writing’s on the walls. *New Phytol.* 2008, *178*, 473–485.

34. Buzała Kamila, P.; Kalinowska, H.; Przybysz, P.; Małachowska, E. Conversion of various types of lignocellulosic biomass to fermentable sugars using kraft pulping and enzymatic hydrolysis. *Wood Sci. Technol.* 2017, *51*, 873–885.
35. Jin, Y.S.; Lee, T.H.; Choi, Y.D.; Ryu, Y.W.; Seo, J.H. Conversion of xylose to ethanol by recombinant Saccharomyces cerevisiae containing genes for xylose reductase and xylitol dehydrogenase from Pichia stipitis. *J. Microbiol. Biotechnol.* 2000, *10*, 564–567.
36. Jönsson, L.J.; Martín, C. Pretreatment of lignocellulose: Formation of inhibitory by-products and strategies for minimizing their effects. *Bioresour. Technol.* 2016.
37. Alvira, P.; Tomás-Pejó, E.; Ballesteros, M.; Negro, M.J. Pretreatment technologies for an efficient bioethanol production process based on enzymatic hydrolysis: A review. *Bioresour. Technol.* 2010.
38. Jönsson, L.J.; Martín, C. Pretreatment of lignocellulose: Formation of inhibitory by-products and strategies for minimizing their effects. *Bioresour. Technol.* 2016.
39. Taylor, M.P.; Mulako, I.; Tuffin, M.; Cowan, D. Understanding physiological responses to pre-treatment inhibitors in ethanologenic fermentations. *Biotechnol. J.* 2012, *7*, 1169–1181.
40. Antal, M.J.; Leesomboon, T.; Mok, W.S.; Richards, G.N. Mechanism of formation of 2-furaldehyde from d-xylose. *Carbohydr. Res.* 1991.
41. Antal, M.J.; Mok, W.S.L.; Richards, G.N. Mechanism of formation of 5-(hydroxymethyl)-2-furaldehyde from d-fructose and sucrose. *Carbohydr. Res.* 1990.
42. Radecka, D.; Mukherjee, V.; Mateo, R.Q.; Stojiljkovic, M.; Foulquié-Moreno, M.R.; Thevelein, J.M. Looking beyond Saccharomyces: The potential of non-conventional yeast species for desirable traits in bioethanol fermentation. *FEMS Yeast Res.* 2015, *15*.
43. Harmsen, P.; Huijgen, W.; López, L.; Bakker, R. Literature Review of Physical and Chemical Pretreatment Processes for Lignocellulosic Biomass. *Food Biobased Res.* 2010, 1–49.
44. Ko, J.K.; Ximenes, E.; Kim, Y.; Ladisch, M.R. Adsorption of enzyme onto lignins of liquid hot water pretreated hardwoods. *Biotechnol. Bioeng.* 2015, *112*, 447–456.
45. Chandel, A.A.K.; Silva, S. da; Singh, O.V.O.; Silvério, S.; Singh, O.V.O. Detoxification of lignocellulosic hydrolysates for improved bioethanol production. *Biofuel Prod. …* 2011, *2012*, 989572.
46. Lin, F.M.; Qiao, B.; Yuan, Y.J. Comparative proteomic analysis of tolerance and adaptation of ethanologenic Saccharomyces cerevisiae to furfural, a lignocellulosic inhibitory compound. *Appl. Environ. Microbiol.* 2009, *75*, 3765–3776.
47. Caspeta, L.; Castillo, T.; Nielsen, J. Modifying Yeast Tolerance to Inhibitory Conditions of Ethanol Production Processes. *Front. Bioeng. Biotechnol.* 2015, *3*, 184.
48. Field, S.J.; Ryden, P.; Wilson, D.; James, S. a; Roberts, I.N.; Richardson, D.J.; Waldron, K.W.; Clarke, T. a Identification of furfural resistant strains of Saccharomyces cerevisiae and Saccharomyces paradoxus from a collection of

environmental and industrial isolates. *Biotechnol. Biofuels* 2015, *8*, 33.
49. Martín, C.; Marcet, M.; Almazán, O.; Jönsson, L.J. Adaptation of a recombinant xylose-utilizing Saccharomyces cerevisiae strain to a sugarcane bagasse hydrolysate with high content of fermentation inhibitors. *Bioresour. Technol.* 2007, *98*, 1767–1773.
50. Zhou, Q.; Liu, Z.L.; Ning, K.; Wang, A.; Zeng, X.; Xu, J. Genomic and transcriptome analyses reveal that MAPK- and phosphatidylinositol-signaling pathways mediate tolerance to 5-hydroxymethyl-2-furaldehyde for industrial yeast Saccharomyces cerevisiae. *Sci. Rep.* 2014, *4*, 6556.
51. Park, S.E.; Koo, H.M.; Park, Y.K.; Park, S.M.; Park, J.C.; Lee, O.K.; Park, Y.C.; Seo, J.H. Expression of aldehyde dehydrogenase 6 reduces inhibitory effect of furan derivatives on cell growth and ethanol production in Saccharomyces cerevisiae. *Bioresour. Technol.* 2011, *102*, 6033–6038.
52. Xiao, H.; Zhao, H. Genome-wide RNAi screen reveals the E3 SUMO-protein ligase gene SIZ1 as a novel determinant of furfural tolerance in Saccharomyces cerevisiae. *Biotechnol. Biofuels* 2014, *7*, 78.
53. Kwon, Y.J.; Ma, A.Z.; Li, Q.; Wang, F.; Zhuang, G.Q.; Liu, C.Z. Effect of lignocellulosic inhibitory compounds on growth and ethanol fermentation of newly-isolated thermotolerant Issatchenkia orientalis. *Bioresour. Technol.* 2011, *102*, 8099–8104.
54. Tenreiro, S.; Nunes, P. a; Viegas, C. a; Neves, M.S.; Teixeira, M.C.; Cabral, M.G.; Sá-Correia, I. AQR1 gene (ORF YNL065w) encodes a plasma membrane transporter of the major facilitator superfamily that confers resistance to short-chain monocarboxylic acids and quinidine in Saccharomyces cerevisiae. *Biochem. Biophys. Res. Commun.* 2002, *292*, 741–748.
55. Sainio, T.; Turku, I.; Heinonen, J. Adsorptive removal of fermentation inhibitors from concentrated acid hydrolyzates of lignocellulosic biomass. *Bioresour. Technol.* 2011, *102*, 6048–6057.
56. Kim, S.K.; Park, D.H.; Song, S.H.; Wee, Y.J.; Jeong, G.T. Effect of fermentation inhibitors in the presence and absence of activated charcoal on the growth of Saccharomyces cerevisiae. *Bioprocess Biosyst. Eng.* 2013, *36*, 659–666.
57. Feldman, D.; Kowbel, D.J.; Glass, N.L.; Yarden, O.; Hadar, Y. Detoxification of 5-hydroxymethylfurfural by the Pleurotus ostreatus lignolytic enzymes aryl alcohol oxidase and dehydrogenase. *Biotechnol. Biofuels* 2015, *8*, 63.
58. Parawira, W.; Tekere, M. Biotechnological strategies to overcome inhibitors in lignocellulose hydrolysates for ethanol production: review. *Crit. Rev. Biotechnol.* 2011, *31*, 20–31.
59. Kim, S.R.; Ha, S.J.; Wei, N.; Oh, E.J.; Jin, Y.S. Simultaneous co-fermentation of mixed sugars: A promising strategy for producing cellulosic ethanol. *Trends Biotechnol.* 2012, *30*, 274–282.
60. Shetty, R.P.; Endy, D.; Knight, T.F. Engineering BioBrick vectors from BioBrick parts. *J. Biol. Eng.* 2008, *2*, 5.
61. Çakar, Z.P.; Turanli-Yildiz, B.; Alkim, C.; Yilmaz, Ü. Evolutionary engineering of Saccharomyces cerevisiae for improved industrially important properties. *FEMS Yeast Res.* 2012, *12*, 171–182.
62. Salusjarvi, L.; Kaunisto, S.; Holmstrom, S.; Vehkomaki, M.L.; Koivuranta, K.; Pitkanen, J.P.; Ruohonen, L. Overexpression of NADH-dependent fumarate

reductase improves d-xylose fermentation in recombinant Saccharomyces cerevisiae. *J. Ind. Microbiol. Biotechnol.* 2013, *40*, 1383–1392.
63. Lindberg, L.; Santos, A.X.S.; Riezman, H.; Olsson, L.; Bettiga, M. Lipidomic Profiling of Saccharomyces cerevisiae and Zygosaccharomyces bailii Reveals Critical Changes in Lipid Composition in Response to Acetic Acid Stress. *PLoS One* 2013, *8*.
64. Dragosits, M.; Mattanovich, D. Adaptive laboratory evolution--principles and applications for biotechnology. *Microb. Cell Fact.* 2013, *12*, 64.
65. Mozhayskiy, V.; Tagkopoulos, I. Microbial evolution in vivo and in silico: Methods and applications. *Integr. Biol. (United Kingdom)* 2013.
66. Oud, B.; Van Maris, A.J.A.; Daran, J.M.; Pronk, J.T. Genome-wide analytical approaches for reverse metabolic engineering of industrially relevant phenotypes in yeast. *FEMS Yeast Res.* 2012.
67. Caspeta, L.; Chen, Y.; Ghiaci, P.; Feizi, A.; Baskov, S.; Hallström, B.M.; Petranovic, D.; Nielsen, J. Altered sterol composition renders yeast thermotolerant. *Science (80-.).* 2014.
68. Rancati, G.; Pavelka, N.; Fleharty, B.; Noll, A.; Trimble, R.; Walton, K.; Perera, A.; Staehling-Hampton, K.; Seidel, C.W.; Li, R. Aneuploidy Underlies Rapid Adaptive Evolution of Yeast Cells Deprived of a Conserved Cytokinesis Motor. *Cell* 2008, *135*, 879–893.
69. Tang, Y.C.; Amon, A. Gene copy-number alterations: A cost-benefit analysis. *Cell* 2013, *152*, 394–405.
70. Hawk, J.D.; Stefanovic, L.; Boyer, J.C.; Petes, T.D.; Farber, R. a Variation in efficiency of DNA mismatch repair at different sites in the yeast genome. *Proc. Natl. Acad. Sci. U. S. A.* 2005, *102*, 8639–8643.
71. Park, C.; Qian, W.; Zhang, J. Genomic evidence for elevated mutation rates in highly expressed genes. *EMBO Rep.* 2012, *13*, 1123–9.
72. Sniegowski, P.D.; Gerrish, P.J. Beneficial mutations and the dynamics of adaptation in asexual populations. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 2010, *365*, 1255–1263.
73. Sauer, U. Evolutionary engineering of industrially important microbial phenotypes. *Adv. Biochem. Eng. Biotechnol.* 2001, *73*, 129–169.
74. Steenwyk, J.L.; Rokas, A. Copy number variation in fungi and its implications for wine yeast genetic diversity and adaptation. *Front. Microbiol.* 2018.
75. Gresham, D.; Desai, M.M.; Tucker, C.M.; Jenq, H.T.; Pai, D.A.; Ward, A.; DeSevo, C.G.; Botstein, D.; Dunham, M.J. The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. *PLoS Genet.* 2008.
76. Hastings, P.J.; Lupski, J.R.; Rosenberg, S.M.; Ira, G. Mechanisms of change in gene copy number. *Nat. Rev. Genet.* 2009.
77. Slack, A.; Thornton, P.C.; Magner, D.B.; Rosenberg, S.M.; Hastings, P.J. On the mechanism of gene amplification induced under stress in Escherichia coli. *PLoS Genet.* 2006.
78. Aguilera, A.; García-Muse, T. Causes of Genome Instability. *Annu. Rev. Genet.* 2013.
79. Hamperl, S.; Cimprich, K.A. Conflict Resolution in the Genome: How Transcription and Replication Make It Work. *Cell* 2016.

80. Stanley, D.; Fraser, S.; Chambers, P.J.; Rogers, P.; Stanley, G.A. Generation and characterisation of stable ethanol-tolerant mutants of Saccharomyces cerevisiae. *J. Ind. Microbiol. Biotechnol.* 2010, *37*, 139–149.
81. Aarnio, T.H.; Suihko, M.L.; Kauppinen, V.S. Isolation of acetic acid-tolerant Baker's yeast variants in a turbidostat. *Appl. Biochem. Biotechnol.* 1991.
82. Almario, M.P.; Reyes, L.H.; Kao, K.C. Evolutionary engineering of Saccharomyces cerevisiae for enhanced tolerance to hydrolysates of lignocellulosic biomass. *Biotechnol. Bioeng.* 2013.
83. Ho, P.W.; Swinnen, S.; Duitama, J.; Nevoigt, E. The sole introduction of two single-point mutations establishes glycerol utilization in Saccharomyces cerevisiae CEN.PK derivatives. *Biotechnol. Biofuels* 2017.
84. Farwick, A.; Bruder, S.; Schadeweg, V.; Oreb, M.; Boles, E. Engineering of yeast hexose transporters to transport D-xylose without inhibition by D-glucose. *Proc. Natl. Acad. Sci. U. S. A.* 2014, *111*, 5159–64.
85. Kurosawa, K.; Laser, J.; Sinskey, A.J. Tolerance and adaptive evolution of triacylglycerol-producing Rhodococcus opacus to lignocellulose-derived inhibitors. *Biotechnol. Biofuels* 2015, *8*, 76.
86. Li, H.; Alper, H.S. Enabling xylose utilization in Yarrowia lipolytica for lipid production. *Biotechnol. J.* 2016.
87. Bonturi, N.; Crucello, A.; Viana, A.J.C.; Miranda, E.A. Microbial oil production in sugarcane bagasse hemicellulosic hydrolysate without nutrient supplementation by a Rhodosporidium toruloides adapted strain. *Process Biochem.* 2017.
88. Caspeta, L.; Nielsen, J. Thermotolerant yeast strains adapted by laboratory evolution show trade-off at ancestral temperatures and preadaptation to other stresses. *MBio* 2015, *6*.
89. Strucko, T.; Zirngibl, K.; Pereira, F.; Kafkia, E.; Mohamed, E.T.; Rettel, M.; Stein, F.; Feist, A.M.; Jouhten, P.; Patil, K.R.; et al. Laboratory evolution reveals regulatory and metabolic trade-offs of glycerol utilization in Saccharomyces cerevisiae. *Metab. Eng.* 2018.
90. Wenger, J.W.; Piotrowski, J.; Nagarajan, S.; Chiotti, K.; Sherlock, G.; Rosenzweig, F. Hunger artists: Yeast adapted to carbon limitation show trade-offs under carbon sufficiency. *PLoS Genet.* 2011.
91. Wouter Wisselink, H.; Toirkens, M.J.; Wu, Q.; Pronk, J.T.; Van Maris, A.J.A. Novel evolutionary engineering approach for accelerated utilization of glucose, xylose, and arabinose mixtures by engineered Saccharomyces cerevisiae strains. *Appl. Environ. Microbiol.* 2009.
92. Bligh, E.G.; Dyer, W.J. A rapid method of total lipid extraction and purification. *Can. J. Biochem. Physiol.* 1959, *37*, 911–917.
93. Capus, A.; Monnerat, M.; Ribeiro, L.C.; de Souza, W.; Martins, J.L.; Sant'Anna, C. Application of high-content image analysis for quantitatively estimating lipid accumulation in oleaginous yeasts with potential for use in biodiesel production. *Bioresour. Technol.* 2016, *203*, 309–317.
94. Manirakiza, P.; Covaci, A.; Schepens, P. Comparative Study on Total Lipid Determination using Soxhlet, Roese-Gottlieb, Bligh & Dyer, and Modified Bligh & Dyer Extraction Methods. *J. Food Compos. Anal.* 2001, *14*, 93–100.
95. De la Hoz Siegler, H.; Ayidzoe, W.; Ben-Zvi, A.; Burrell, R.E.; McCaffrey, W.C.

Improving the reliability of fluorescence-based neutral lipid content measurements in microalgal cultures. *Algal Res.* 2012, *1*, 176–184.
96. Doan, T.T.Y.; Obbard, J.P. Improved Nile Red staining of Nannochloropsis sp. *J. Appl. Phycol.* 2011, *23*, 895–901.
97. Rumin, J.; Bonnefond, H.; Saint-Jean, B.; Rouxel, C.; Sciandra, A.; Bernard, O.; Cadoret, J.-P.; Bougaran, G. The use of fluorescent Nile red and BODIPY for lipid measurement in microalgae. *Biotechnol. Biofuels* 2015, *8*, 42.
98. Kimura, K.; Yamaoka, M.; Kamisaka, Y. Rapid estimation of lipids in oleaginous fungi and yeasts using Nile red fluorescence. *J. Microbiol. Methods* 2004.
99. Sitepu, I.R.; Ignatia, L.; Franz, A.K.; Wong, D.M.; Faulina, S.A.; Tsui, M.; Kanti, A.; Boundy-Mills, K. An improved high-throughput Nile red fluorescence assay for estimating intracellular lipids in a variety of yeast species. *J. Microbiol. Methods* 2012, *91*, 321–328.
100. Rostron, K.A.; Lawrence, C.L. Nile red staining of neutral lipids in yeast. In *Methods in Molecular Biology*; 2017; Vol. 1560, pp. 219–229.
101. Fowler, S.; Brown, W.; Warfel, J.; Greenspan, P. Use of nile red for the rapid in situ quantitation of lipids on thin-layer chromatograms. *J Lipid Res* 1987.
102. Pick, U.; Rachutin-Zalogin, T. Kinetic anomalies in the interactions of Nile red with microalgae. *J. Microbiol. Methods* 2012, *88*, 189–196.
103. Bertozzini, E.; Galluzzi, L.; Penna, A.; Magnani, M. Application of the standard addition method for the absolute quantification of neutral lipids in microalgae using Nile red. *J. Microbiol. Methods* 2011, *87*, 17–23.
104. Orr, V.; Rehmann, L. Improvement of the Nile Red fluorescence assay for determination of total lipid content in microalgae independent of chlorophyll content. *J. Appl. Phycol.* 2014, *27*, 2181–2189.
105. Raschke, D.; Knorr, D. Rapid monitoring of cell size, vitality and lipid droplet development in the oleaginous yeast Waltomyces lipofer. *J. Microbiol. Methods* 2009, *79*, 178–183.
106. Da Silva, T.L.; Feijão, D.; Reis, A. Using multi-parameter flow cytometry to monitor the yeast rhodotorula glutinis CCMI 145 batch growth and oil production towards biodiesel. *Appl. Biochem. Biotechnol.* 2010, *162*, 2166–2176.
107. da Silva, T.L.; Santos, C.A.; Reis, A. Multi-parameter flow cytometry as a tool to monitor heterotrophic microalgal batch fermentations for oil production towards biodiesel. *Biotechnol. Bioprocess Eng.* 2009, *14*, 330–337.
108. Smith, A.E.; Zhang, Z.; Thomas, C.R.; Moxham, K.E.; Middelberg, a P. The mechanical properties of Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. U. S. A.* 2000, *97*, 9871–4.
109. Notman, R.; Noro, M.; O'Malley, B.; Anwar, J. Molecular basis for dimethylsulfoxide (DMSO) action on lipid membranes. *J. Am. Chem. Soc.* 2006, *128*, 13982–13983.
110. Wang, Z.; Zhuge, J.; Fang, H.; Prior, B.A. Glycerol production by microbial fermentation: A review. *Biotechnol. Adv.* 2001.
111. Vongsvivut, J.; Heraud, P.; Gupta, A.; Puri, M.; McNaughton, D.; Barrow, C.J. FTIR microspectroscopy for rapid screening and monitoring of polyunsaturated fatty acid production in commercially valuable marine

yeasts and protists. *Analyst* 2013, *138*, 6016–31.
112. Ami, D.; Posteri, R.; Mereghetti, P.; Porro, D.; Doglia, S.M.; Branduardi, P. Fourier transform infrared spectroscopy as a method to study lipid accumulation in oleaginous yeasts. *Biotechnol. Biofuels* 2014, *7*, 12.
113. Huang, Y.Y.; Beal, C.M.; Cai, W.W.; Ruoff, R.S.; Terentjev, E.M. Micro-Raman spectroscopy of algae: Composition analysis and fluorescence background behavior. *Biotechnol. Bioeng.* 2010, *105*, 889–898.
114. Oro, L.; Ciani, M.; Comitini, F. Antimicrobial activity of Metschnikowia pulcherrima on wine yeasts. *J. Appl. Microbiol.* 2014.
115. Robledo-Leal, E.; Rodriguez-Padilla, C.; Rivera-Morales, L.G. Identification and susceptibility of clinical isolates of Candida spp. to killer toxins. *Brazilian J. Biol.* 2018, *78*.
116. Kluyver, J.; Van der Walt, J.P.; van Triet, a. J. Pulcherrimin, The Pigment of Candida Pulcherrima. *Proc . N.* 1953, *39*, 583–593.
117. Chantasuban, T.; Santomauro, F.; Gore-Lloyd, D.; Parsons, S.; Henk, D.; Scott, R.J.; Chuck, C. Elevated production of the aromatic fragrance molecule, 2-phenylethanol, using Metschnikowia pulcherrima through both de novo and ex novo conversion in batch and continuous modes. *J. Chem. Technol. Biotechnol.* 2018.
118. Sitepu, I.; Selby, T.; Lin, T.; Zhu, S.; Boundy-Mills, K. Carbon source utilization and inhibitor tolerance of 45 oleaginous yeast species. *J. Ind. Microbiol. Biotechnol.* 2014, *41*, 1061–1070.
119. Long, Z.; Santomauro, F.; Fan, J.; Macquarrie, D.; Clark, J.; Chuck, C.J.; Budarin, V. Fast microwave-assisted acidolysis: a new biorefinery approach for the zero-waste utilisation of lignocellulosic biomass to produce high quality lignin and fermentable saccharides. *Faraday Discuss.* 2017.
120. Gorter de Vries, A.R.; Pronk, J.T.; Daran, J.M.G. Industrial relevance of chromosomal copy number variation in Saccharomyces yeasts. *Appl. Environ. Microbiol.* 2017.
121. Veitia, R.A.; Bottani, S.; Birchler, J.A. Cellular reactions to gene dosage imbalance: genomic, transcriptomic and proteomic effects. *Trends Genet.* 2008.
122. Safonova, Y.; Bankevich, A.; Pevzner, P.A. DIPSPADES: Assembler for highly polymorphic diploid genomes. In Proceedings of the Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); 2014.
123. Stanke, M.; Morgenstern, B. AUGUSTUS: A web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* 2005.
124. Byrne, K.P.; Wolfe, K.H. The Yeast Gene Order Browser: Combining curated homology and syntenic context reveals gene fate in polyploid species. *Genome Res.* 2005.
125. Li, H.; Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 2010.
126. Wolfgert, H.; Manmun, Y.M.; Kuchler, K. The Yeast Pdr15p ATP-binding Cassette (ABC) Protein Is a General Stress Response Factor Implicated in Cellular Detoxification. *J. Biol. Chem.* 2004.
127. Larochelle, M.; Drouin, S.; Robert, F.; Turcotte, B. Oxidative Stress-Activated

Zinc Cluster Protein Stb5 Has Dual Activator/Repressor Functions Required for Pentose Phosphate Pathway Regulation and NADPH Production. *Mol. Cell. Biol.* 2006.
128. Sasano, Y.; Watanabe, D.; Ukibe, K.; Inai, T.; Ohtsu, I.; Shimoi, H.; Takagi, H. Overexpression of the yeast transcription activator Msn2 confers furfural resistance and increases the initial fermentation rate in ethanol production. *J. Biosci. Bioeng.* 2012.
129. Alriksson, B.; Horváth, I.S.; Jönsson, L.J. Overexpression of Saccharomyces cerevisiae transcription factor and multidrug resistance genes conveys enhanced resistance to lignocellulose-derived fermentation inhibitors. *Process Biochem.* 2010.
130. Sato, T.K.; Liu, T.; Parreiras, L.S.; Williams, D.L.; Wohlbach, D.J.; Bice, B.D.; Ong, I.M.; Breuer, R.J.; Qin, L.; Busalacchi, D.; et al. Harnessing genetic diversity in saccharomyces cerevisiae for fermentation of xylose in hydrolysates of alkaline hydrogen peroxide-pretreated biomass. *Appl. Environ. Microbiol.* 2014, *80*, 540–554.
131. Westman, J.O.; Taherzadeh, M.J.; Franzén, C.J. Inhibitor tolerance and flocculation of a yeast strain suitable for second generation bioethanol production. *Electron. J. Biotechnol.* 2012.
132. Petersson, A.; Almeida, J.R.M.; Modig, T.; Karhumaa, K.; Hahn-Hägerdal, B.; Gorwa-Grauslund, M.F.; Lidén, G. A 5-hydroxymethyl furfural reducing enzyme encoded by the Saccharomyces cerevisiae ADH6 gene conveys HMF tolerance. *Yeast* 2006, *23*, 455–464.
133. Causton, H.C.; Ren, B.; Koh, S.S.; Harbison, C.T.; Kanin, E.; Jennings, E.G.; Lee, T.I.; True, H.L.; Lander, E.S.; Young, R.A. Remodeling of Yeast Genome Expression in Response to Environmental Changes. *Mol. Biol. Cell* 2001.
134. Sadeh, A.; Movshovich, N.; Volokh, M.; Gheber, L.; Aharoni, A. Fine-tuning of the Msn2/4-mediated yeast stress responses as revealed by systematic deletion of Msn2/4 partners. *Mol. Biol. Cell* 2011.
135. Mira, N.P.; Becker, J.D.; Sá-Correia, I. Genomic Expression Program Involving the Haa1p-Regulon in *Saccharomyces cerevisiae* Response to Acetic Acid. *Omi. A J. Integr. Biol.* 2010.
136. Kim, J.E.; Jang, I.S.; Sung, B.H.; Kim, S.C.; Lee, J.Y. Rerouting of NADPH synthetic pathways for increased protopanaxadiol production in Saccharomyces cerevisiae. *Sci. Rep.* 2018.
137. Schweizer, M.; Roberts, L.M.; Höltke, H.J.; Takabayashi, K.; Höllerer, E.; Hoffmann, B.; Müller, G.; Köttig, H.; Schweizer, E. The pentafunctional FAS1 gene of yeast: its nucleotide sequence and order of the catalytic domains. *MGG Mol. Gen. Genet.* 1986.
138. Wenz, P.; Schwank, S.; Hoja, U.; Schüller, H.J. A downstream regulatory element located within the coding sequence mediates autoregulated expression of the yeast fatty acid synthase gene FAS2 by the FAS1 gene product. *Nucleic Acids Res.* 2001.
139. Mira, N.P.; Palma, M.; Guerreiro, J.F.; Sá-Correia, I. Genome-wide identification of Saccharomyces cerevisiae genes required for tolerance to acetic acid. *Microb. Cell Fact.* 2010.
140. Adeboye, P.T.; Bettiga, M.; Olsson, L. ALD5, PAD1, ATF1 and ATF2 facilitate

the catabolism of coniferyl aldehyde, ferulic acid and p-coumaric acid in Saccharomyces cerevisiae. *Sci. Rep.* 2017.
141. Ploier, B.; Scharwey, M.; Koch, B.; Schmidt, C.; Schatte, J.; Rechberger, G.; Kollroser, M.; Hermetter, A.; Daum, G. Screening for hydrolytic enzymes reveals Ayr1p as a novel triacylglycerol lipase in Saccharomyces cerevisiae. *J. Biol. Chem.* 2013.
142. Lee, Y.; Nasution, O.; Lee, Y.M.; Kim, E.; Choi, W.; Kim, W. Overexpression of PMA1 enhances tolerance to various types of stress and constitutively activates the SAPK pathways in Saccharomyces cerevisiae. *Appl. Microbiol. Biotechnol.* 2017.
143. Rodriguez-Peña, J.M.; Cid, V.J.; Arroyo, J.; Nombela, C. The YGR194c (XKS1) gene encodes the xylulokinase from the budding yeast Saccharomyces cerevisiae. *FEMS Microbiol. Lett.* 1998.
144. Ouyang, L.; Holland, P.; Lu, H.; Bergenholm, D.; Nielsen, J. Integrated analysis of the yeast NADPH-regulator Stb5 reveals distinct differences in NADPH requirements and regulation in different states of yeast metabolism. *FEMS Yeast Res.* 2018.